

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: April 11, 2002, 08:53:27 ; Search time 13.3 Seconds
(without alignments)
1251.566 Million cell updates/sec

Title: Perfect score: 2447
Sequence: 1 MGENDPPAVEAPFSSFRSLFG. TRVTSFLUDWIHQMRDQLKT 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

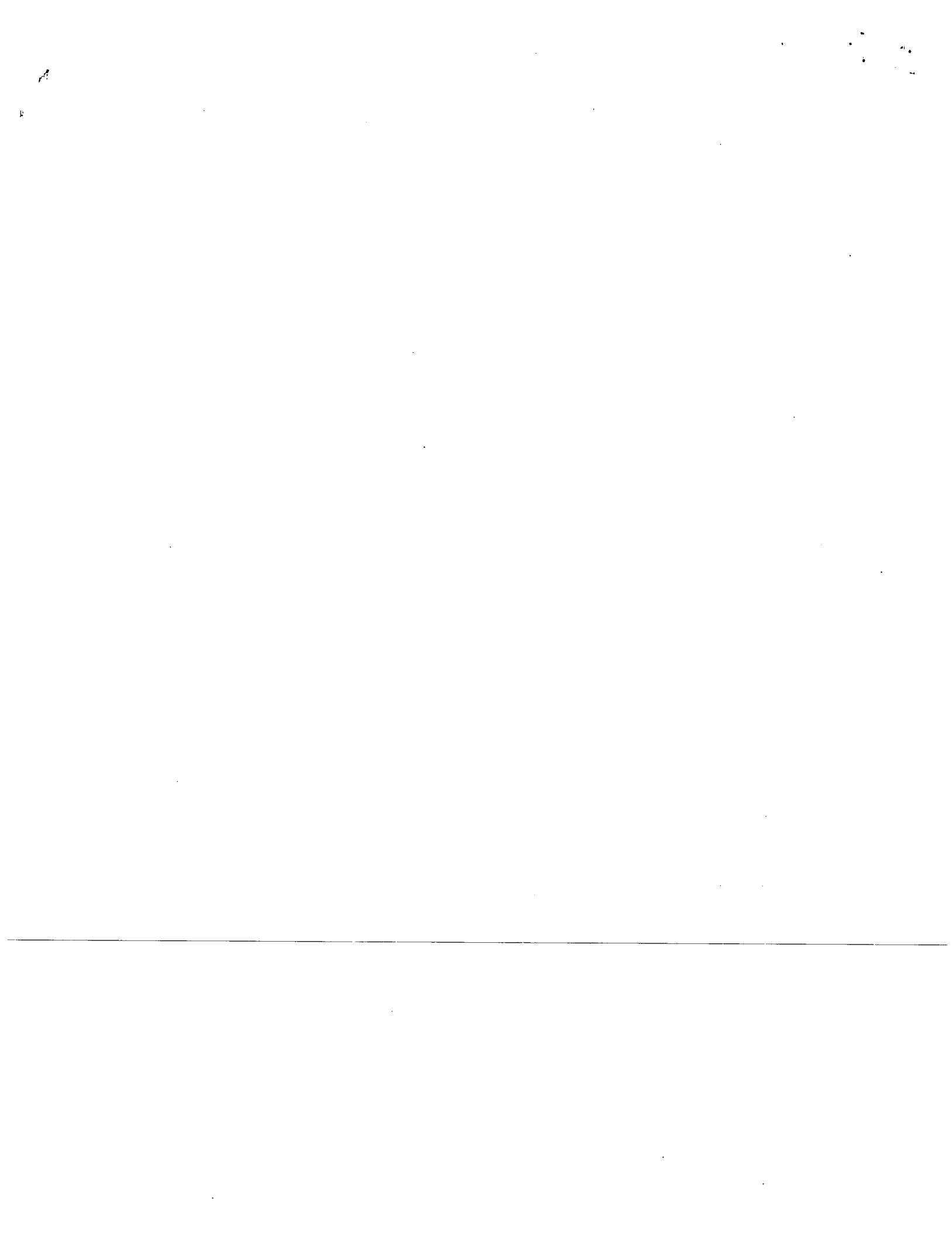
SUMMARIES

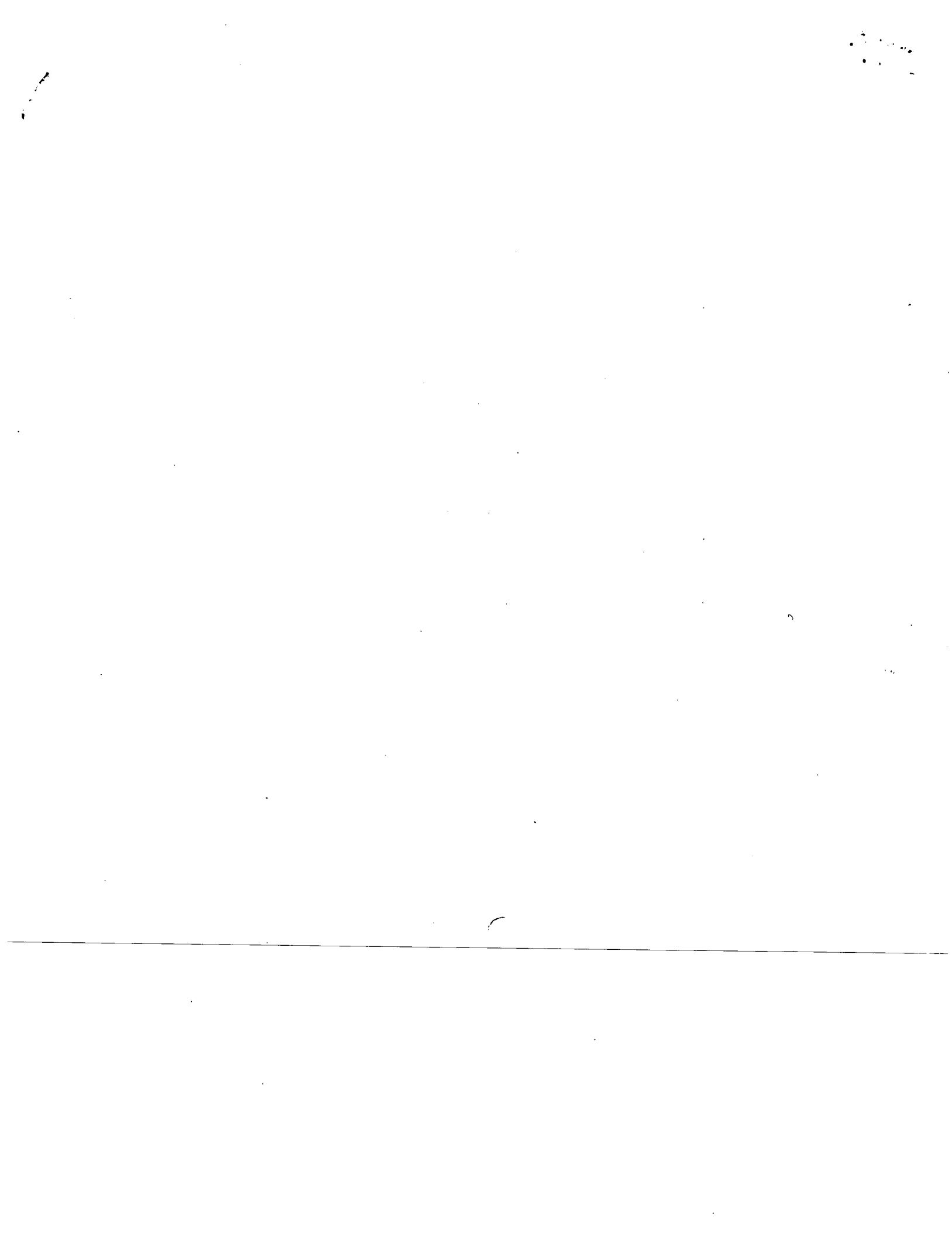
Result No.	Score	Query	Length	DB ID	Description
1	2447	100.0	454	1 TMS3_HUMAN	P57727 homo sapien
2	865.5	35.4	490	1 TMS2_MOUSE	0915393 mus musculus
3	864.5	35.3	492	1 TMS2_HUMAN	09er04 homo sapien
4	704	28.8	455	1 TMS5_MOUSE	09h333 homo sapien
5	696	28.4	457	1 TMS5_HUMAN	09h331 homo sapien
6	664.5	27.2	417	1 HEPS_HUMAN	P05931 Homo sapien
7	661.5	27.0	1069	1 ENTK_MOUSE	P97455 mus musculus
8	660	27.0	437	1 TMS4_HUMAN	09ns4 homo sapien
9	660	27.0	1035	1 ENTK_BOVIN	P98072 Bos taurus
10	657	26.8	1019	1 ENTK_HUMAN	P98073 Homo sapien
11	654	26.7	1034	1 ENTK_PIG	P98074 Sus scrofa
12	650	26.6	416	1 HEPS_MOUSE	035653 mus musculus
13	645	26.4	416	1 HEPS_RAT	005511 rattus norvegicus
14	570	23.3	638	1 KAL_RAT	P14712 Rattus norvegicus
15	561	22.9	638	1 KAL_MOUSE	P26822 mus musculus
16	540	22.1	638	1 KAL_HUMAN	P03952 Homo sapien
17	533	21.8	855	1 SP14_MOUSE	09Y566 homo sapien
18	532	21.7	324	1 TEST_MOUSE	09jnj7 mus musculus
19	526	21.5	790	1 PLMN_FIG	P06877 sus scrofa
20	523.5	21.1	671	1 NETR_MOUSE	008762 mus musculus
21	520.5	21.3	338	1 PLMN_HORSE	P80010 equus caballus
22	520	21.3	1042	1 CORI_HUMAN	Q9y955 homo sapien
23	512.5	20.9	855	1 SP14_MOUSE	P56677 mus musculus
24	508.5	20.8	314	1 TEST_HUMAN	Q9y600 homo sapien
25	507	20.7	812	1 PLMN_MOUSE	P20918 mus musculus
26	507	20.7	437	1 ACRO_RAT	P29933 rattus norvegicus
27	506	20.7	1113	1 CORI_MOUSE	Q9-319 mus musculus
28	504.5	20.6	810	1 PLMN_HUMAN	P00747 homo sapien
29	503.5	20.6	273	1 MCT7_MOUSE	P02844 mus musculus
30	503.5	20.6	436	1 ACRO_MOUSE	P25718 mus musculus
31	503	20.6	343	1 PLMN_SHEEP	P81266 ovis aries
32	501.5	20.5	810	1 PLMN_ERIEC	Q29485 erinaceus europaeus
33	501.5	20.5	875	1 METR_HUMAN	P56730 homo sapien

ALIGNMENTS

P15944 canis familiaris
P10323 homo sapiens
Q9es87 ratmus norvegicus
P48038 oryctolagus cuniculus
P06868 bos taurus
P12545 mactaca mulatta
P50342 meriones unguiculatus
P80009 canis familiaris
Q9esdi mus musculus
P03951 homo sapiens
P21845 mus musculus
P50343 rattus norvegicus

RESULT 1					
ID	TMS3_HUMAN	STANDARD;	PRT;	454 AA.	
AC	P57727;				
DT	20-AUG-2001 (Rel. 40, created)				
DT	20-AUG-2001 (Rel. 40, Last sequence update)				
DE	TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE TADG-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).				
DE	TADG-12 (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).				
GN	TMRSS3 OR TADG12 OR ECHS1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Primates; Catarhini; Hominoidea; Homo.				
OX					
RN					
RP					
RX					
RA	Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N., Wang Y., Parmley T.H., O'Brien T.J.;				
RA	"Ovarian tumor cells express a novel multi-domain cell surface serine protease."				
RL	Biochim. Biophys. Acta 1502:337-350(2000).				
RX					
RA	SEQUENCE FROM N.A. (ISOFORMS A, B/C AND D), AND VARIANT IIE-53.				
RX	[2]				
RA	SCOTT H.S., KUDOL J., WATTENHOFER M., SHIBUYA K., BERRY A., CHRST. R., GUIPONI M., WANG J., KAWASAKI K., ASAKAWA S., MINOSHIMA S., YOUNIS F., MEIDI S.O., RADHAKRISHNA U., PAPASAVVAS M.P., GEHRIG C., ROSSER C., KOROSTISHERSKY M., GAL A., SHIMIZU N., BONNE-TAMIR B., ANTONARAKIS S.E.;				
RA	ANTONARAKIS S.E.;				
RA	Insertions of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness.;				
RT	RT				
RT	protease causing both congenital and childhood onset autosomal recessive deafness.;				
RL	Nat. Genet. 27:59-63 (2001).				
CC					
CC	- - FUNCTION: PROBABLE PROTEASE.				
CC	- - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).				
CC	- - ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND TRUNCATED TADG-12V. ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	- - TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED TADG-12 IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.				
CC	- - DISEASE: DEFECTS IN TMRSS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL NEUROSENSE CHILDHOOD ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10.				
CC	- - SIMILARITY: BELONGS TO PEPTIDE FAMILY SI; ALSO KNOWN AS THE TRYPSIN FAMILY.				
CC	- - SIMILARITY: CONTAINS 1 SRCR DOMAIN.				
CC	- - SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.				
CC	CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				





PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US1274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillian KJ;
 PI KJavvin IU, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Sheldon DL, Stewart TA, Tunas D, Williams PM, Wood WI;
 DR WPI; 2000-611443/58.
 XX
 Novel PRO polypeptides and polynucleotides used in detection methods, to target bioactive molecules to specific cells, and to modulate cellular activities -
 XX
 PS Claim 12; Fig 28; 636pp; English.
 XX
 C78458 to C78599 represent polynucleotide and EST (expressed sequence tag) sequences which encode secreted or transmembrane PRO polypeptides. The PRO polynucleotides and polypeptides have cytostatic activity. The polynucleotides and polypeptides can be used for detecting the presence of PRO polypeptides in samples, for linking bioactive molecules to cells and for modulating biological activities of cells, using the polypeptides for specific targeting. The polypeptide targeting can be used to kill the target cells, e.g. for the treatment of cancers. The polypeptide pairs provide specific targeting of bioactive molecules to cells. C78600 to C78987 represent PCR primers and probes used in the isolation of the PRO polynucleotide sequences.
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 XX
 Sequence 453 AA;

Query Match 99.9%; Score 2441; DB 21; Length 453;
 Best Local Similarity 99.8%; Pred. No. 2.6e-188; Indels 0; Gaps 0;
 Matches 452; Conservative 1; Mismatches 0;
 QY 1 MGENDPAPAVEAPTSSFRSLFGLDDIKISVAPADAAQIISLPLRKFPIVIGIALLI 60
 Db 1 mgendpapaveaptssfrslfglldikisvapadaaqilslplrifpilwigiali 60
 61 LATAGLGLHFDGSGKRYCRSSFFCIELTARCDCGSDCKDGEDPYRCVUGGONAVLQF 120
 61 laaiglighfdgskrycrssffcicieiarcdgsvckdgedpyrcvuggqavilqf 120
 121 TAASIKTMCSDDWKGHYANVACAGIGEFSYVSSNLYRSVSLQESQFREERFVSDIDLPPDK 180
 121 taasiktmcsddwkghyanvacagigefsvssnlyrsvslqesqfrefervsdidlppdk 180
 Db 181 VTAHLISVYVREGASGHVTLQTAGHRRGYSRIVGNGSISQSPWQASQFOGH 240
 QY 181 VTAHLISVYVREGASGHVTLQTAGHRRGYSRIVGNGSISQSPWQASQFOGH 240
 Db 181 vtahlisyvregasghvtlqtaghrrgysrivgngsilsqpwqasqfqqgh 240
 QY 241 LCGGSVITPLWITAAHCVTDLYLQPKSWTQVGLVSLDQAPPSHLVEKIVHSKYKRR 300
 241 lcggsvitplwitaahcvttdlylpkswtqvgylsldopapshivekivhskypkr 300
 Db 301 LGNDIAALMKLAGPLTENEMTQPVCLPNESENPDPCKVCMWSGGWATEDEGDDASPVLNHA 360
 301 lgndiaalmkagpltenemtqpvclpneenpdpckvcmwsggwategeddaspvlnha 360
 Db 301 lgndiaalmkagpltenemtqpvclpneenpdpckvcmwsggwategeddaspvlnha 360
 361 VPLISNKTCNRDYGKGISSMCAAGYLGGPSQGDSGGPVCOERLUWKVAGTSF 420
 241 lcgsvitplwitaahcvttdlylpkswtqvgylsldopapshivekivhskypkr 300
 Db 361 vplisnkcnhrdvyggiispsmleagyltgvdscqgdsqgplvcqerrlwkvagtsf 420
 QY 421 GIGCAEVNPKPGYTRTSFLDMIHEQMERDJKT 453
 Db 421 giccaevnpgytrvtsfldmihqmerdjk 453

RESULT 2
 Y41694
 ID Y41694 standard; Protein; 452 AA.
 XX
 Y41694;
 XX
 DT 07-DEC-1999 (first entry)
 XX
 DE Human PRO382 protein sequence.
 XX
 Human: PRO; EST; expressed sequence tag; PCR primer; hybridisation; probe; blood coagulation disorder; cancer; cellular adhesion disorder; secreted protein; transmembrane protein.
 KW OS Homo sapiens.
 XX PN WO9946281-A2.
 XX PD 16-SEP-1999.
 XX PR 08-MAR-1999; 99WO-US05028.
 XX PR 10-MAR-1998; 98US-0077450.
 PR 11-MAR-1998; 98US-0077632.
 PR 11-MAR-1998; 98US-0077641.
 PR 11-MAR-1998; 98US-0077649.
 PR 12-MAR-1998; 98US-0077791.
 PR 13-MAR-1998; 98US-0078004.
 PR 17-MAR-1998; 98US-0040220.
 PR 20-MAR-1998; 98US-0078866.
 PR 20-MAR-1998; 98US-0078910.
 PR 20-MAR-1998; 98US-0078936.
 PR 20-MAR-1998; 98US-0078939.
 PR 25-MAR-1998; 98US-0079294.
 PR 26-MAR-1998; 98US-0079656.
 PR 27-MAR-1998; 98US-0079663.
 PR 27-MAR-1998; 98US-0079669.
 PR 27-MAR-1998; 98US-0079766.
 PR 27-MAR-1998; 98US-0079766.
 PR 30-MAR-1998; 98US-0079923.
 PR 31-MAR-1998; 98US-0080105.
 PR 31-MAR-1998; 98US-0080107.
 PR 31-MAR-1998; 98US-0080165.
 PR 31-MAR-1998; 98US-0080194.
 PR 01-APR-1998; 98US-0080327.
 PR 01-APR-1998; 98US-0080328.
 PR 01-APR-1998; 98US-0080333.
 PR 01-APR-1998; 98US-0080334.
 PR 08-APR-1998; 98US-0081049.
 PR 08-APR-1998; 98US-0081070.
 PR 08-APR-1998; 98US-0081152.
 PR 09-APR-1998; 98US-0081195.
 PR 09-APR-1998; 98US-0081203.
 PR 09-APR-1998; 98US-0081229.
 PR 15-APR-1998; 98US-0081817.
 PR 15-APR-1998; 98US-0081838.
 PR 15-APR-1998; 98US-0081955.
 PR 15-APR-1998; 98US-0082568.
 PR 21-APR-1998; 98US-0082569.
 PR 21-APR-1998; 98US-0082700.
 PR 22-APR-1998; 98US-0082704.
 PR 22-APR-1998; 98US-0082804.
 PR 23-APR-1998; 98US-0082867.
 PR 23-APR-1998; 98US-0083336.
 PR 27-APR-1998; 98US-0083392.
 PR 28-APR-1998; 98US-0083392.
 PR 29-APR-1998; 98US-0083395.

CoreCore version 4.5 Copyright (c) 1993 - 2000 Compugen Ltd.						
protein - protein search, using sw model						April 11, 2002, 08:52:32 ; Search time 15.91 Seconds (without alignments)
un on:						642.143 Million cell updates/sec
Title: US-09-846-512-12						Effect score: 2447
Sequence: 1 MGENDPPAVEAEPFSRSLFG..... TRVTSFLDWIHEQMERDLKT 454						Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5						Search space: 212252 seqs, 22303292 residues
Total number of hits satisfying chosen parameters: 212252						Post-processing: Minimum Match 0% Maximum Match 100%
Listing first 45 summaries						
Database : Issued_Patents_AA:*						
1: /eqn2_6/ptodata/2/1aa/5A_COMB.pep:*						Score: 2375
2: 2195.5 97.1 454 4 US-09-518-046-2						Length: 2195.5
3: 1354 55.3 294 4 US-09-261-416-2						Sequence: 1, Appli
4: /eqn2_6/ptodata/2/1aa/5A_COMB.pep:*						Sequence 2, Appli
5: /eqn2_6/ptodata/2/1aa/5A_COMB.pep:*						Sequence 4, Appli
6: /eqn2_6/ptodata/2/1aa/5A_COMB.pep:*						Sequence 23, Appli
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.						
SUMMARIES						
Result No.	Score	Query Match	Length	DB	ID	Description
1	2375	97.1	454	4	US-09-518-046-2	Sequence 2, Appli
2	2195.5	89.7	455	4	US-09-261-416-2	Sequence 2, Appli
3	1354	55.3	294	4	US-09-518-046-4	Sequence 4, Appli
4	875.5	35.8	492	4	US-09-342-749-2	Sequence 2, Appli
5	831.5	34.8	157	4	US-09-518-046-23	Sequence 23, Appli
6	686.5	28.1	283	3	US-09-807-151-1	Sequence 1, Appli
7	660.5	27.0	435	4	US-09-008-271A-6	Sequence 6, Appli
8	660	27.0	798	1	US-09-200-900A-2	Sequence 2, Appli
9	660	27.0	798	1	PCT-US94-00616-2	Sequence 5, Appli
10	655.5	26.8	416	2	US-09-000-846-2	Sequence 2, Appli
11	576	23.5	256	2	US-09-227-337-3	Sequence 3, Appli
12	571	23.3	255	4	US-09-944-483-67	Sequence 67, Appli
13	570	23.3	638	2	US-09-611-151-3	Sequence 3, Appli
14	565.5	23.1	418	1	US-09-508-448C-25	Sequence 2, Appli
15	533	21.8	232	1	US-09-508-448C-19	Sequence 19, Appli
16	533	21.8	855	2	US-09-027-337-2	Sequence 2, Appli
17	520.5	21.6	248	4	US-09-944-483-63	Sequence 63, Appli
18	527	21.5	235	3	US-09-807-151-3	Sequence 3, Appli
19	523	21.4	98	4	US-09-518-046-17	Sequence 17, Appli
20	518	21.2	235	4	US-09-944-483-65	Sequence 65, Appli
21	509.5	20.8	546	6	520340-6	Patent No. 520340
22	508.5	20.8	314	4	US-09-008-271A-3	Sequence 3, Appli
23	507.5	20.7	812	1	US-09-248-629A-1	Sequence 1, Appli
24	507.5	20.7	812	1	US-09-451-932-1	Sequence 1, Appli
25	507.5	20.7	812	1	US-09-452-260-1	Sequence 1, Appli
26	507.5	20.7	812	1	US-09-326-785-1	Sequence 1, Appli
27	507.5	20.7	812	2	US-09-612-788-1	Sequence 1, Appli

RESULT 3
US-09-518-046-4
; Sequence 4, Application US/09518046
; Patent No. 629463
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
TITLE OF INVENTION: Transmembrane Serine Protease Overexpressed
TITLE OF INVENTION: in Ovarian Carcinoma and Uses Thereof
FILE REFERENCE: D6192CIP
CURRENT APPLICATION NUMBER: US/09/518,046
CURRENT FILING DATE: 2000-03-02
; EARLIER APPLICATION NUMBER: 09/261,416
EARLIER FILING DATE: 1999-03-03
; SEQ ID NO 4
; SEQ ID NO 4
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
FEATURE:
; OTHER INFORMATION: complete amino acid sequence of TRDG-12
; OTHER INFORMATION: variant protein
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
NUMBER OF SEQ ID NOS: 14

SEQ ID NO 2
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo sapiens
FEATURE:
; OTHER INFORMATION: Amino acid sequence of TRDG-12 encoded by nucleotides
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663

Query Match 89.7%; Score 2195.5; DB 4; Length 455;
Best Local Similarity 91.5%; Pred. No. 6e-207; Gaps 4;
Matches 421; Conservative 5; Mismatches 23; Indels 11; Gaps 4;

Query 1 MGENDPPAVEAPPSFRSLFLGDDLKKISPVAPADAVAAQIILSLPLKFF-PIVIGITAL 59
1 MGENDPPAVEAPPSFRSLFLGDDLKKISPVAPADAVAAQIILSLPLKFF-PIVIGITAL 59

Query 60 ILALAIIGIIGIHDCCSGKVRCRSSFKCIELIARCDGVSDCKDQGDEDEYRCVRVYGGQNAVQV 119
60 ILALAIIGIIGIHDCCSGKVRCRSSFKCIELIARCDGVSDCKDQGDEDEYRCVRVYGGQNAVQV 119

Query 61 ILALAIIGIIGIHDCCSGKVRCRSSFKCIELIARCDGVSDCKDQGDEDEYRCVRVYGGQNAVQV 120
61 ILALAIIGIIGIHDCCSGKVRCRSSFKCIELIARCDGVSDCKDQGDEDEYRCVRVYGGQNAVQV 120

Query 120 FIASWKIMCSDDWKGHAYANVACQALGFPSPYVSSDNRLVSSLEGQFREFVSDIDLPPD 179
121 FIASWKIMCSDDWKGHAYANVACQALGFPSPYVSSDNRLVSSLEGQFREFVSDIDLPPD 180

Query 180 KVITALHHSVYVREGCASGHVWTLQCTAGHRRGKGSYSSRIVGGNMSLISQWPMQASLQFQGY 239
Db 181 KVITALHHSVYVREGCASGHVWTLQCTAGHRRGKGSYSSRIVGGNMSLISQWPMQASLQFQGY 239

Query 240 HLCGGSVITPLWITTAACVYDYLPLKSTWQGLVSL-----LQNPARSH 285
Db 241 HLCGGSVITPLWITTAACVYDYLPLKSTWQGLVSL-----LQNPARSH 285

Query 120 FIASWKIMCSDDWKGHAYANVACQALGFPSPYVSSDNRLVSSLEGQFREFVSDIDLPPD 179
Db 121 FIASWKIMCSDDWKGHAYANVACQALGFPSPYVSSDNRLVSSLEGQFREFVSDIDLPPD 180

Query 180 KVITALHHSVYVREGCASGHVWTLQCTAGHRRGKGSYSSRIVGGNMSLISQWPMQASLQFQGY 239
Db 181 KVITALHHSVYVREGCASGHVWTLQCTAGHRRGKGSYSSRIVGGNMSLISQWPMQASLQFQGY 239

Query 240 HLCGGSVITPLWITTAACVYDYLPLKSTWQGLVSL-----LQNPARSH 285
Db 241 HLCGGSVITPLWITTAACVYDYLPLKSTWQGLVSL-----LQNPARSH 285

RESULT 4
US-09-342-749-2
; Sequence 2, Application US/09342749
; Patent No. 6166194
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavagian, Sean V.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: IMPRSS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: 144 to 1511 of Sequence 1
; Patent No. 6291663
NUMBER OF SEQ ID NOS: 14

Query 300 RGGNDIAALMKLAGPLTNEMLQPCVCLNSENEMPDGKWCWGGWATEDGADASPLNH 359
301 RGGNDIAALMKLAGPLTNEMLQPCVCLNSENEMPDGKWCWGGWATEDGADASPLNH 359

Query 360 AAPVLTSNPKICHHRDVGIIISPSMILCAGYLGQV-----SCQGSGGGPLVQCRERLWK 414
360 AAPVLTSNPKICHHRDVGIIISPSMILCAGYLGQV-----SCQGSGGGPLVQCRERLWK 414

Query 415 LYCATSGIGCAEVNPKGVYTRTSFLDWIHQMERDIFT 454
416 LYCATSGIGCAEVNPKGVYTRTSFLDWIHQMERDIFT 455

Query 416 LYCATSGIGCAEVNPKGVYTRTSFLDWIHQMERDIFT 455

Db 250 TSWGSGCAKARPGVIGNVWFTDWIYRQMRAD 282

RESULT 7

US-09-008-271A-6

; Sequence 6, Application US/09008271A

Patent No. 6203979

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Hillman, Jennifer L.

Yue, Henry

Guegler, Karl J.

Corley, Neil C.

Tang, Tom Y.

Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271A

FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheila

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 435 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: COLIN013

CLONE: 1337018

SEQUENCE DESCRIPTION: SEQ ID NO: 6 :

US-09-008-271A-6

Query Match 27.0%; Score 660.5; DB 4; Length 435; Best Local Similarity 35.4%; Pred. No. 1.6e-56; Matches 157; Conservative 73; Mismatches 150; Indels 63; Gaps 17;

Qy 44 IPIKFFPPIVIGIITALALA----IGLIGHFDCSGKRYRORSSFKC--TELIIR-- 91

Qy 22 IPMETFRKGIVIRIALLSLASIIIVVLLKILD--KY---YELCGQPLHFFRKQL 74

Qy 92 CDGVSDCKDGDEYRC-----VVGQNAVLQFVAA--SWKTMQSDDWKGHYA 138

Db 75 CDEGELDPLGEDEEHCVKSFPPGAVAVRLSDRSIQLVIDSATGWWPSACFDDNFEAL 134

Qy 139 NVACAQLGFPSSVSSDNIRVSSLEGQFEEFVSDHLPDDKWTALHHHSVYREG--CA 195

Qy 135 ETACROMGYSS-----KPTFRAVEIGPDQDLDVVEITENSOELMRNNSGPCL 182

Qy 196 SCHVWILQCTAGCHRRGQYSSRIVGGNNMSLLSOWPQASLQFOGGYHLCGGSVITPLWITA 255

Db 183 SSSLVSLHCLAGGESL-KTPRIVGGEFAASVDSWPWQVSQYDQHVGGSILDPHWLTA 241

Qy 256 ARCV--YDLYLPKSMITQVGLVSLDNPAPSHLVKIV--YHSYKPKLGLNDIALMK 309

Db 242 AHCFRKHDFV--NNWKRAGSDKL--GSFPPLSLAVAKIITEFNPWY-PK--DNDIALMK 293

Qy 310 LAGPLTNEMIQPVCLPNSSEENPDGKVCWTSWGATEDAGDASVSLNHAUPLTSNKI 369

Db 294 LQFPPLTFSGTVPRICLDPFDEELTPATPLWITGWFPTKQNGKMSDILLQASVQIDSTR 353

Qy 370 CNHRDVIGGITSPSMLCAGYLTTGGVSDQGDSGGPLVQCRRLWKLVAGASFGIGCAEV 429

Db 354 CNADDAVQGEVIEKMMCAGIPBGGVMDQGSGGPMYQSDQ-WHWVGIVSWGIGGGPS 412

Qy 430 KPGVYTRVTSFLDWHQMERDL 452

Db 413 TPGVYTKVSYAYLNWLNWKAEL 435

RESULT 8

US-08-200-900A-2

; Sequence 2, Application US/08200900A

; Patent No. 5665566

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc. - Legal Affairs

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: MA

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (BPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/200,900A

FILING DATE: 23-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Meinert, Maureen C.

REGISTRATION NUMBER: 31,544

REFERENCE/DOCKET NUMBER: GI 5201-FWC

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170 X8574

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 798 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-200-900A-2

Query Match 27.0%; Score 660; DB 1; Length 798; Best Local Similarity 36.9%; Pred. No. 4.3e-56; Matches 146; Conservative 70; Mismatches 146; Indels 34; Gaps 15;

Qy 66 GLGIGHFDC-SGKRYRCSRFKCLEIARCDGVSCKDQEDERCVRV--GGQNAVLQFV 120

Db 415 GLGIPPEPKEDNEQCKDG-ECIPLVNLCDGFPHCKGDSDEAHCVRLFNGTTRDSSGLVQFR 473

Qy 121 TAASWKTMCSDDWKGHYANVACQALGFPSSVSSDNIRVSSLEGQFEEFVSDHLPDDK 180

Db 474 IOSIWHVACAEWNTQISDQVQQLGIGT--GNSVSPFTSPGG--GPVWNL----- 521

Qy 181 VITALHHHSVY--REGCASGHVVTLQCT--ACGHR--RGYSSRIVGGNNMSLLSQWPWQAS 233

RESULT 9
 PCT-US94-00616-2
 ; Sequence 2, Application PC/US9400616
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
 ; NUMBER OF SEQUENCES: 33
 ; COMPUTER READABLE FORM:
 ; COMPUTER: IBM PC compatible
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/00616
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 798 amino acids
 ; TYPE: amin acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US94-00616-2

Query Match 27.0%; Score 660; DB 5; Length 798;
 Best Local Similarity 36.9%; Pred. No. 4.3e-56;
 Matches 146; Conservative 70; Mismatches 146; Indels 34; Gaps 15;

QY 66 GLGIRHDFC-SGKVRGRSSFKIELTARCDQVSDCKDGEDEBYRCRVV---GGONAVLOVE 120
 Db 415 GLGFEPCKEDNFQCKDG-ECIPLVNLCDGFPKHDGSDEAHCVRLFNGTTDSSGLVQFR 473
 QY 121 TAASKTMTSDDWKGHYANVACAGQLGFPSSVSSNLRLVSLEGQFREERVSIDHLPPDK 180
 Db 474 IOSTIWHVACAEWNWTOQISDVOQCOLQGLGT--GNNSVPTFSTGG--GPYVNLN----- 521
 QY 181 VTAHLHSVYV--RECCASGHVWTLOCT--ACGHR--RGSSRVGGNNMSLSQLQWPWQAS 233
 Db 522 -TAPNGSLTTPSQQCLEDSSLILQCNVYRSGKQKLYTQVSPKIVGGSRSREGWPWVA 580
 CQ 234 LQFQOYHLCGGSVITPLWITTAACVYDYL-PHSWTIYGL-VSLIDNP-ABSHLYEK 289
 Db 581 LYFDQQVQGCCASLNSRDWLVSAAHCYGRNMEPSKWKAVLGLHMASNLSPQETRLIDQ 640
 QY 290 IVVHSKYKPKRLGNDIALMLKLAGPLPFTNEMIOPVCLPNSSENFDPGKVWTSGMATEG 349
 Db 641 IVINPHYNKRKNNDIAMMILEMKVNTDYIOPCLPEENQVFPPGRICSIAGNGAL-TY 699
 QY 350 AGDASPVLNHAAPLISNKCNHR-DVYGGIISMSLCAQYLTCGVDSQGDGSGPLVQ 408
 Db 700 OGSTADVLQADVPULSNERKQOQMPNEY--ITENNCAGYEAGVGDSCQGDGSGPLMCQ 757
 Db 758 ENRRLLAGVTSFGYOCALPNRPGVYARVPRFTENI 793

RESULT 10
 US-09-000-846-2
 ; Sequence 2, Application US/09000846
 ; GENERAL INFORMATION:
 ; APPLICANT: WU, QINGYU
 ; APPLICANT: SADLER, JASPER
 ; TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH
 ; TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSE: MILLIN, WHITE, ZELANO & BRANIGAN, P.C.
 ; STREET: 2200 CLARENDON BLVD. SUITE 1400
 ; CITY: ARLINGTON
 ; STATE: VA
 ; COUNTRY: US
 ; ZIP: 22201
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0. Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/000,846
 ; FILING DATE: 30-DEC-1997
 ; CLASSIFICATION: 800
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/9866,058
 ; FILING DATE: 30-MAY-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LEBOVITZ, RICHARD M.
 ; REGISTRATION NUMBER: 37, 067
 ; REFERENCE/DOCKET NUMBER: BBRX 65P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-243-6333
 ; TELEFAX: 703-243-6410
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 416 amino acids
 ; TYPE: amin acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-000-846-2

Query Match 26.8%; Score 655; DB 2; Length 416;
 Best Local Similarity 38.0%; Pred. No. 4.6e-56;
 Matches 143; Conservative 54; Mismatches 136; Indels 43; Gaps 10;

QY 100 DGEDFYRCVRYGGQNQAVLQVF--TAASKTMTSDDWKGHYANVACAGQLGFPSSVSSDNIR 157
 Db 46 DQEPLXQ-VOLSPGDSRLAVFDKFTGTVRLLCSRSNRVAGLGCEEMGFFRLAHSELD 104
 QY 158 VSSLEGQFREREFVSD-----HLLPDKVTLAHHSVYVRECCASGHVWTLOCTAGH 209
 Db 105 VRTAGANGTSSGFFCVDDEGGIQLRQLL--DVISVC-----DCPGRGFLTATCQDCGR 154
 Qy 210 RRGSSRIVGNNMSLSQLQWQASLQFOQYHLCGGSVITPLWITTAACVYDYLPRK-- 266
 Db 155 RKLVPDRINGQDGSLSLGRWQVSLRDPYTHLQCGSLUSGDDWVLTAAIC----FPRNR 209
 QY 267 ---SWTIQVGLVSLIDNPNAPSHLVEKIVHSKYKPR---LGNDIALMLKLAGPLTN 317
 Db 210 VLSRMVRFAGAVARTSPHAWQLGQVAVIHYGGYLPFRDPTIDENSNDIALVHLSLPLT 269
 Qy 318 EMIOVCLNSEEWFPGKWCWTSWGATEDGADASPVTLHAAVPLISNKICNHRDYG 377

RESULT 11
US-09-027-337-3
Sequence 3, Application US/09027337B
; Patent No. 597616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotoshi.
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: DE0664
; CURRENT APPLICATION NUMBER: US/09/027-337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 3
; LENGTH: 256
; TYPE: PRM
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; US-09-027-337-3

Query Match 23.5%; Score 576; DB 2; Length 256;
Best Local Similarity 45.1%; Pred. No. 1; 4e-48;
Matches 114; Conservative 37; Mismatches 74; Indels 28; Gaps 7;

QY 216 RTGGNNMSLLSOWPQASLQFQGGYHLCGGSVTPTLWITAACVYDLYLPK-----SWT 269
Db 1 RIVGGRBTLGRKWPQWVSLRVDGAHLCGGSLISGDWLTAAHC-----FPERNRVLSWR 55

QY 270 IQVGLVSLLDNPAFPHL---VETKIVVSKYKRR-----LGNDIALMKLAGPLTENEMI 320
Db 56 VFAAGAVA---QASPHGLQGVDAVWYHGGYLFRDPNSEENSDIALVHLSSPLPLTEYI 112

QY 321 QPVCLPNSSENFPGKWCWTSQGATEDAGDASPVLNHAAPLISNKICNRDVGGII 380
Db 113 QPVCLPAAQGQLVDGKICTVTWGNQ-YGQOAGVILQEARVPIITSNDVNCAGDFYGNQI 171

QY 381 SPMSMCAGYLRTGGVDSQGDSGGPLVQCE---RRWKLVAGTSFGIGCAEVNKGPGYTR 436
Db 172 KPMFECAGYPEGGIDACQGDSSGPFPVCEDSISRTPRWRLLCGIVSMWTGCAALQKPGVYTK 231

QY 437 VTFSLDWHQE 449
Db 232 VSDFREWFQAIK 244

RESULT 12
US-08-944-083-67
Sequence 67, Application US/08944483
; Patent No. 622456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

RESULT 13
US-08-681-151-3
Sequence 3, Application US/08681151
; Patent No. 586637
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Braxton, Scott Michael

TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
TITLE OF INVENTION: OF THE PROSTATE
NUMBER OF SEQIDENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
; INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 255 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456
; US-08-944-483-67

Query Match 23.3%; Score 571; DB 4; Length 255;
Best Local Similarity 44.8%; Pred. No. 4; 4e-48;
Matches 113; Conservative 37; Mismatches 74; Indels 28; Gaps 7;

QY 217 IVGGNNMSLLSOWPQASLQFQGYHLCGGSVTPTLWITAACVYDLYLPK-----SWT 270
Db 1 IVGGRBTLGRKWPQWVSLRVDGAHLCGGSLISGDWLTAAHC-----FPERNRVLSWR 55

QY 271 QVGLVSLLDNPAFPHL---VETKIVVSKYKRR-----LGNDIALMKLAGPLTENEMI 321
Db 56 FAGAVA---QASPHGLQGVDAVWYHGGYLFRDPNSEENSDIALVHLSSPLPLTEYI 112

QY 322 PVCLPNSSENFPGKWCWTSQGATEDAGDASPVLNHAAPLISNKICNRDVGGII 381
Db 113 PVCLPAAQGQLVDGKICTVTWGNQ-YGQOAGVILQEARVPIITSNDVNCAGDFYGNQI 171

QY 382 PSMLCAGYLRTGGVDSQGDSGGPLVQCE---RRWKLVAGTSFGIGCAEVNKGPGYTR 437
Db 172 KPMFECAGYPEGGIDACQGDSSGPFPVCEDSISRTPRWRLLCGIVSMWTGCAALQKPGVYTK 231

APPLICANT: Goli, Surya
 TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: US
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSEQ Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/681,151
 FILING DATE: Herewith
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0074US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-555-0555
 TELEFAX: 415-841-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 638 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Peptide
 IMMEDIATE SOURCE:
 LIBRARY: GENBANK
 CLONE: 205011
 8-681-151-3

RESULT 14
US 08-508-448C-25
; Sequence 25, Application US/08508448C
; Patent No. 5804410
; GENERAL INFORMATION:

Search completed: April 11, 2002, 08:53:24
Job time: 52 sec

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:53:27 ; Search time 13.3 Seconds

(without alignments)

1251.566 Million cell updates/sec

Title:

US-09-846-512-12

Perfect score:

2447

Sequence: 1 MGENDPPAVEAPFSRSFLG.TRVTSFLDWIHEQMERDLKT 454

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt-39;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2447	100.0	454	1 TMS3_HUMAN	P57727 homo sapien
2	865.5	35.4	490	1 TMS2_MOUSE	09j98 mus musculus
3	864.5	35.3	492	1 TMS2_HUMAN	015393 homo sapien
4	704	28.8	455	1 TMS5_MOUSE	09er0 mus musculus
5	696	28.4	457	1 TMS5_HUMAN	09hs33 homo sapien
6	664.5	27.2	417	1 HEPS_HUMAN	P05981 homo sapien
7	661.5	27.0	1069	1 HEPS_MOUSE	P97435 mus musculus
8	660.5	27.0	437	1 TMS4_HUMAN	09nrs4 homo sapien
9	660	27.0	1035	1 ENTK_BOVIN	P98072 bos taurus
10	656	26.8	1019	1 ENTK_HUMAN	P98073 homo sapien
11	654	26.7	1034	1 ENTK_PIG	P98074 sus scrofa
12	650	26.6	416	1 HEPS_MOUSE	035453 mus musculus
13	645	26.4	416	1 HEPS RAT	005511 rattus norvegicus
14	570	23.3	638	1 KAL1_RAT	P14272 rattus norvegicus
15	561	22.9	638	1 KAL_MOUSE	P26262 mus musculus
16	540	22.1	638	1 KAL_HUMAN	P03952 homo sapien
17	533	21.8	855	1 ST14_HUMAN	Q9Y5Y6 homo sapien
18	532	21.7	324	1 TEST_MOUSE	Q9jh7 mus musculus
19	526	21.5	791	1 PLMN_PIG	P05677 sus scrofa
20	523.5	21.4	761	1 NEFR_MOUSE	008762 mus musculus
21	520.5	21.3	338	1 PLMN_HOUSE	P80010 equus caballus
22	520	21.3	1042	1 CORI_HUMAN	Q9Y545 homo sapien
23	512.5	20.9	855	1 ST14_MOUSE	P56677 mus musculus
24	508.5	20.8	314	1 TEST_HUMAN	Q9Y600 homo sapien
25	507.5	20.7	812	1 PLMN_MOUSE	P20918 mus musculus
26	507	20.7	437	1 ACRO_RAT	P29293 rattus norvegicus
27	506	20.7	1113	1 CORI_MOUSE	Q9z319 mus musculus
28	504.5	20.6	810	1 PLMN_HUMAN	P00747 homo sapien
29	503.5	20.6	273	1 ST17_MOUSE	Q92844 mus musculus
30	503.5	20.6	436	1 ACRO_MOUSE	P23578 mus musculus
31	503	20.6	343	1 PLMN_SHEEP	P81286 ovis aries
32	501.5	20.5	810	1 PLMN_ERIEU	Q92485 erinaceus europaeus
33	501.5	20.5	875	1 NETR_HUMAN	P56730 homo sapien

ALIGNMENTS

RESULT 1

TMS3_HUMAN

ID TMS3_HUMAN STANDARD; PRT; 454 AA.

AC P57727; DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.7) (SERINE PROTEASE TADG-12) (TUMOR ASSOCIATED DIFFERENTIALLY-EXPRESSED GENE-12 PROTEIN).

DE TMPRSS3 OR TADG12 OR ECHOS1.

GN Homo sapiens (Human)

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OC NCBI_TaxID:9606; OX NCBL_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).

RC TISSUE=Ovarian carcinoma;

RX MEDLINE=20521358; PubMed=11068177; MEDLINE=20578749; PubMed=11137999;

RA Underwood L.J., Shitomas K., Tanimoto H., Beard J.B., Schneider E.N., Ra Wang Y., Parmley T.H., O'Brien T.J.; "Ovarian tumor cells express a novel multi-domain cell surface serine protease.", Biochim. Biophys. Acta 1602:337-350(2000).

RA Scott H.S., Kudoh J., Wackenhoffer M., Shibusawa K., Berry A., Chrast R., RA Qipponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S., Younis F., Mehdi S.O., Radhakrishna U., Papasavvas M.P., Gehrig C., RA Antonarakis S.E.; "Insertion of beta-gatellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness.", Nat. Genet. 27:59-63(2001).

RA - FUNCTION: PROTEOLYTIC PROTEASE.

CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).

CC - ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND TRUNCATED/TADG-12V; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC - TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.

CC - DISEASE: DEFECTS IN TMPRSS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DFNB8 AND DFNB10.

CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY 51; ALSO KNOWN AS THE TRYPSIN FAMILY.

CC - SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

P15944 canis familiaris

P10323 homo sapiens

Q9es87 rattus norvegicus

P48038 oryctolagus cuniculus

P06868 bos taurus

P12545 macaca mulatta

P50342 meriones unicolor

P80009 canis familiaris

Q9esdl mus musculus

P03951 homo sapiens

P21845 mus musculus

P50343 ratitornis norvegicus

DR	EMBL; AF201380; AAQ37012.1;	OY	1	MGENDPPAVAPFSFRSLFGLDPLIKTSPVAPDAAQIISLPLKFFPILVIGIALLI
DR	EMBL; AB038157; BAB20077.1; -.	1		
DR	EMBL; AB038158; BAB20078.1; -.	1		
DR	EMBL; AB038159; BAB2079.1; -.	1		
DR	EMBL; AB038160; BAB20080.1; -.	1		
DR	MIM: 605111; -.	61	LATAIGIIGIHEODSGKVRGRSSKTCIELJARDGVSQDKGDEYRVRVSQNAVQF	
DR	MIM: 605167; -.	61		
DR	InterPro; IPR001314; CHYMOTRYPSIN.	120		
DR	InterPro; IPR002172; LDL_Recept_A.	120		
DR	InterPro; IPR001100; SRCR.	180		
DR	InterPro; IPR001254; TRYPSIN.	180		
DR	Pfam; PF00057; LDL_recept_a; 1.	121	TAASWKTMCSDMKGHYANAVACQOLGPSPYSVSDNLVSLQGOFREFFVSIDHLLPDK	
DR	Pfam; PF000530; SRCR; 1.	121		
DR	Pfam; PF00089; TRYPSIN; 1.	180		
DR	SMART; SM00192; LDL; 1.	180		
DR	SMART; SM00202; SR; 1.	180		
DR	SMART; SM00200; TRYPSIN_SPEC; 1.	180		
DR	PROSITE; PS001209; LDLR_A; 1.	180		
DR	PROSITE; PS00068; LDLR_B; 1.	180		
DR	PROSITE; PS00420; SRCR; 1; FALSE_NEG.	240		
DR	PROSITE; PS50287; SRCR; 2; 1.	240		
DR	PROSITE; PS50240; TRYPSIN_DOMAIN; 1.	300		
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.	300		
DR	KW	360		
FT	Hydrolase; Serine protease; Transmembrane; Signal-anchor; Deafness;	360		
FT	Alternative splicing; Polymorphism.	360		
FT	DOMAIN	360		
FT	TRANSMEM	360		
FT	70	454		
FT	72	108		
FT	109	205		
FT	217	454		
FT	257	257		
FT	304	304		
FT	401	401		
FT	216	217		
FT	73	85		
FT	79	98		
FT	92	107		
FT	129	194		
FT	142	204		
FT	207	324		
FT	242	258		
FT	338	407		
FT	370	386		
FT	397	425		
FT	CARBOHYDROXY	221		
FT	1	127		
FT	454	454		
FT	VARSPLIC	318		
FT	VARSPLIC	261	293	
FT	VARSPLIC	294	454	
FT	VARIANT	53	53	
FT	/FTid=VAR_010781.			
FT	LKEFPITV->FEVFSQSSL (IN REF. 1).			
FT	A->T (IN REF. 1).			
FT	MISSING (IN REF. 1).			
FT	MISSING (IN REF. 1).			
FT	MISSING (IN TRUNCATED ISOFORM).			
FT	V->I.			
FT	CONFFLICT	46	54	
FT	CONFFLICT	90	90	
FT	CONFFLICT	350	350	
FT	CONFFLICT	369	395	
FT	SEQUENCE	427	E->D (IN REF. 1).	
FT	SEQUENCE	454	AA->D (IN REF. 1).	
FT	SEQUENCE	49404	AA; MW; 57EC3678FD6AFF CRC64;	
FT	Query	100.0%	Score 2447; DB 1; Length 454;	
FT	Match	100.0%	Best Local Similarity 100.0%; Pred. No. 6; ge-201; Mismatches 0; Indels 0; Gaps 0;	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC

DR HSSP; P00761; LARS.
 DR MSD; MGI:1354381; Impress2.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR01254; Trypsin.
 DR Pfam; PF00057; LDL_recept_a; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMONTRYPSIN.
 DR SMART; SM00192; LDLa; 1.
 DR SMART; SM00302; SR; 1.
 DR SMART; SM00320; TRYP_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS0287; SRCR_2; 1.
 DR PROSITE; PS50240; TRPSIN_DOM; 1.
 DR PROSITE; PS0134; TRPSIN_HIS; 1.
 DR PROSITE; PS0135; TRPSIN_SER; 1.
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 FT DOMAIN 1 83 CTOPLASMIC (POTENTIAL).
 FT TRANSMEM 84 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT - DOMAIN 105 490 (POTENTIAL).
 FT DOMAIN 111 149 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 150 242 LDL-RECEPTOR CLASS A.
 FT DOMAIN 254 490 SERINE PROTEASE.
 FT ACT_SITE 294 294 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 253 254 CLEAVAGE (POTENTIAL).
 FT DISULFID 76 125 BY SIMILARITY.
 FT DISULFID 119 138 BY SIMILARITY.
 FT DISULFID 132 147 BY SIMILARITY.
 FT DISULFID 171 230 BY SIMILARITY.
 FT DISULFID 184 240 BY SIMILARITY.
 FT DISULFID 243 363 BY SIMILARITY.
 FT DISULFID 279 295 BY SIMILARITY.
 FT DISULFID 408 424 BY SIMILARITY.
 FT DISULFID 435 463 BY SIMILARITY.
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 122 122 S -> L (IN REF. 3).
 FT CONFLICT 178 178 S -> N (IN REF. 3).
 FT CONFLICT 320 320 Y -> H (IN REF. 1).
 FT CONFLICT 474 474 D -> D (IN REF. 1).
 SO SEQUENCE 490 AA; 53479 MW; 072B03E4D8A1A9 CRC64;

Query Match 35.4%; Score 865.5%; DB 1; Length 490;
 Best Local Similarity 40.9%; Pred. No. 3.4e-66; Mismatches 146; Index 67; Gaps 14;

QY 28 PVPAPDARAQVQLSLPLKRPPIV'-----IGII-----A 58
 |||:|||:|||:|||:|||:|||:
 31 PVAPNG-----VNLYPAQYIPSPVQYPAQYRPTQOASTSIVIHTPKSSGAPCTSKSKS 83

QY 59 LILALALIGGI-----HFDCS GKYRCRSSRCIELJARCQGVSYOCKGED 103
 |||:|||:|||:|||:|||:|||:|||:|||:
 84 LCLALALGIVLGAAYAVALLWRFWDNSCSTSEMEGSSGTCISSLWMCDCGVAHCPNGED 143

QY 104 EYRCVVRGGNAVLQVFTA-- ASWKTCMSDDWKGHVNACQAQLGFP-SVVSQSDLRLVSS 160
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 144 ENRCVRLYQGQSFILQVYSSQRKAWYVFCQDDWSESYGRACKDMGYKNNYVSSQGIPDOS 203
 CC 161 LEGQFRBEFSIDHLUDDPKYTALHISVYVREGCAGHVVYIQLCAGHHR-GYSSRIVG 219
 |||:|||:|||:|||:|||:|||:|||:
 QY 220 GNMSSLSQWPMQASLOFQGHGKLGGSVITPIWITRAHCYD-LYLPKSTIQVGLV-S 276
 |||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 257 GLNAGSGDWPWQSVLIVQWVQGGSITPWIYTAHCYFEPISGPRYNTAFACTLRS 316
 |||:|||:|||:|||:|||:|||:|||:
 QY 277 LLDPNAPASHVLYKEVHSKKYKRGNDALMKGALPQPLTNEQIOPVCLDNEENFPDSK 336
 |||:|||:|||:|||:|||:|||:
 Db 317 LMFY-GSRHQEVKIVSHPSKTKNDALMKGQPLANFLKREVCLPNPGMLDDQ 375
 |||:|||:|||:|||:
 QY 337 VCWTSGWGAATDAGGAGASPVNHAAPV 396
 |||:|||:|||:|||:|||:|||:
 Db 376 ECWISGNGATYE-KGKTSVDLNAAAMPPLIEPSKCNKSYKVNLLTIPAMICAGFLGSVD 434
 |||:|||:|||:|||:
 QY 397 COGDSEGPVQOERRWKLKGATSGIGCAGEVNKPGVYTHVTSFEDWIKHQ 448
 |||:|||:|||:|||:|||:
 Db 435 COGDSEGPVLUKNGIWNWLTDTSMSGCKALRPGVYGVNTFEDWIKHQ 486
 |||:|||:|||:

RESULT 3

ID TMS2_HUMAN STANDARD; PRT; 492 AA.

AC 015393; DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).
 GN TMRSS2.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Cetartiodata; Homidae; Homo.
 OX NCBI_TaxID=9606;

RA Paoloni-Giacobino A., Chen H., Peisch M.C., Rossier C.,
 RA Antonarakis S.E.,
 RT "Cloning of the TMRSS2 gene, which encodes a novel serine protease
 RT with transmembrane, LDLA, and SRCR domains and maps to 21q22.3.";
 RT Genomics 44:309-320 (1997).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RX PubMed=114763;
 RA Teng D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
 RT "Mutation analyses of 268 candidate genes in human tumor cell lines.",
 RT Genomics 74:352-364 (2001).
 RN [3]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=2110430; PubMed=1169526;
 RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Viikko P.T.;
 RT "Expression of transmembrane serine protease TMRSS2 in mouse and
 RT human tissues.",
 RL J. Pathol. 193:134-140 (2001).
 CC - - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC - - TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO
 CC EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.
 CC - - SIMILARITY: BELONGS TO PEPTIDASE FAMILY 51; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC - - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC - - SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@ist-sib.ch).
DR	InterPro; IPR02172; LDL_recept_A.
DR	InterPro; IPR001190; SRCR.
DR	InterPro; IPR001254; Trypsin.
DR	Pfam; PF00057; ldl_recept_a; 1.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PRO0722; CHYMOTRYPSIN.
DR	SMART; SM0020; SR; 1.
DR	SMART; SM0020; TRYSPC; 1.
DR	PROSITE; PS01209; LDRA_1; 1.
DR	PROSITE; PS5008; LDRA_2; 1.
DR	PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR	PROSITE; PS00487; SRCR_2; 1.
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS0015; TRYPSIN_SER; 1.
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT	DOMAIN 1 84 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 85 105 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
FT	DOMAIN 106 492 EXTRACELLULAR (POTENTIAL).
FT	ACT_SITE 345 345 LDL-RECEPTOR CLASS A.
FT	DOMAIN 150 242 SRCR.
FT	DOMAIN 256 492 SERINE PROTEASE.
FT	ACT_SITE 296 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DOMAIN 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 441 441 CLEAVAGE (POTENTIAL).
FT	SITE 255 256 BY SIMILARITY.
FT	DISUFID 113 126 BY SIMILARITY.
FT	DISUFID 120 139 BY SIMILARITY.
FT	DISUFID 133 148 BY SIMILARITY.
FT	DISUFID 172 231 BY SIMILARITY.
FT	DISUFID 185 241 BY SIMILARITY.
FT	DISUFID 244 365 BY SIMILARITY.
FT	DISUFID 281 297 BY SIMILARITY.
FT	DISUFID 410 426 BY SIMILARITY.
FT	DISUFID 437 465 BY SIMILARITY.
FT	CARBONYD 213 213 N-LINED (GLCNAC . . .) (POTENTIAL).
FT	CARBONYD 249 249 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT	CONFLICT 242 242 L -> I (IN REF. 2).
FT	CONFLICT 329 329 Q -> E (IN REF. 2).
FT	CONFLICT 449 449 N -> K (IN REF. 2).
FT	CONFLICT 491 491 KAN -> RAD (IN REF. 2).
FT	SEQUENCE 492 AA; 53847 MW; TEAAFFDA18609DDA CRC64; SQ
Query	Match 35.3%; Score 864.5; DB 1; Length 492;
Best Local Similarity 45.4%; Pred. No. 4.2e-56;	
Matches 186; Conservative 55; Mismatches 148; Indels 21; Gaps 11;	
QY 52 I V I G I T A L I A L A I G I G I H F -- D O S G K - Y R C R S R F K C I L I A R C D G V S C K D G E D E Y R C 107	
Db 89 I L T G I F L V G A L A A G I L W K I M G K S C S G N S C T C I P N S N W C G V S I C P G G E D E N R C 148	
QY 108 V R V G G O N A L O V F T A - A S W K T M C S D D W K H Y A N V A C A Q I G F P - S V V S S O N L R V S S L E G Q 164	
Db 149 V R V G G P N F I Q M Y M S S O R K S H P V C Q D D W M E N Y G R A C R D M G K N F Y S S O G I V D S G T S 208	
QY 165 F R E E F V S I D H I L P D D K V T A L H S V V Y R E G G A S G H V T L Q T A C G H R R G Y S - S R T V G N M 222	
Db 209 F M K L N T S A G N V - - D T I K K Y H S - - - D A C S S K A V I S L R C I A C G N I N S S R Q S R I V G G E S 261	
QY 223 S L S O W P W Q O S L Q F O G Y H I G G S V I T P L W I T A H C V - Y D Y L P K S W T I Q V G L V - S L D 279	
Db 262 A L P G W P Q W Q S L R H V N V H G G S I T P E W M T A A C V E K P I N P R W H T A P I G L Q F S M E 321	
CC	PRINTS; PRO0722; CHYMOTRYPSIN.
DR	SMART; SM0020; TRYSPC; 1.
DR	PROSITE; PS00240; TRYPSIN_DOM; 1.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
DR	PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR	PROSITE; PS50287; SRCR_2; FALSE_NEG.
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor.
FT	DOMAIN 1 49 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 50 70 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)

FT	DOMAIN	71	(POTENTIAL).
FT	DOMAIN	112	455 EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	218	SRCR.
FT	ACT-SITE	258	SERINE PROTEASE.
FT	ACT-SITE	308	CHARGE RELAY SYSTEM (BY SIMILARITY):
FT	ACT-SITE	405	CHARGE RELAY SYSTEM (BY SIMILARITY):
FT	SITE	217	CHARGE RELAY SYSTEM (BY SIMILARITY):
FT	DISULFID	135	CLEAVAGE (POTENTIAL).
FT	DISULFID	148	BY SIMILARITY.
FT	DISULFID	206	BY SIMILARITY.
FT	DISULFID	328	BY SIMILARITY.
FT	DISULFID	243	BY SIMILARITY.
FT	DISULFID	374	BY SIMILARITY.
FT	DISULFID	491	BY SIMILARITY.
FT	CARBOHYD	163	N-LINKED GLCNAC. (POTENTIAL).
FT	CARBOHYD	170	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	319	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	375	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1	MISSING (IN ISOFORM 2).
FT	VARSPLIC	10	MISSING (IN ISOFORM 3).
FT	VARSPLIC	182	MISSING (IN ISOFORM 1).
FT	VARSPLIC	183	GGLVEEAKWP > MEAQVGLLWV (IN ISOFORM 1).
FT	CONFLICT	325	D > G (IN REF. 1; BAB2077).
SEQUENCE	455 AA;	49632 MN;	5CF31789C6899AA CRC64;

RESULT 9
ENRK_BOVIN STANDARD; PRT; 1035 AA.
ID ENRK_BOVIN
AC P8072;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
GN PRSS7 OR ENRK.
OS Bos taurus (Bovine).
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OC Bovidae; Bos.
NCBI_TAXID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RL Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains.";
PROC. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RN [2]
RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=9404122; PubMed=8226855;
RA Lavallie E.R., Remenault E.A., Diblasio E.A., Ferenz C., Grant K.L., Light A., McCoy J.M.;
RT "Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase.";
RL J. Biol. Chem. 268:23311-23317(1993).
RX SEQUENCE OF 801-827.
RC TISSUE=Intestine;
RX MEDLINE=92189715; PubMed=1799406;
RA Light A., Janska R.;
RT "the amino-terminal sequence of the catalytic subunit of bovine enterokinase";
RL J. Protein Chem. 10:475-480(1991).
RX PROTEIN: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTrypsin AND CARBOXYPEPTIDASE A). IT CATALYZES THE CONVERSION OF TRYPSINogen TO TRYPSIN WHICH IN TURN CATALYZES OTHER PROENZYMES INCLUDING CHYMOTrypsinogen, PROCARBOXYPEPTIDASES, AND PROBLASTASES.
RX -I- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN TRYPSINogen.
CC -I- SUBUNIT: HETERO-DIMER OF A CATALYTIC (LIGHT) CHAIN AND A MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -I- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -I- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS CLEARED BY A TRYPSIN-LIKE PROTEASE.
CC -I- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -I- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -I- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation. The European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

RA	Carboxyhyd	486	486	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	519	519	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	550	550	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	646	646	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	698	698	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	722	722	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	741	741	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	762	762	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	864	864	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	903	903	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONHD	965	965	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	166	192	MISSING (IN SHORT ISOFORM).
FT	CONFFLICT	808	808	R -> Y (IN REF. 3).
SQ	SEQUENCE	1035	AA:	114887 MW: E207970B08296E13 CRC64;
Query Match		27.0%	Score	660; DB 1; Length 1035;
Best Local Similarity		36.9%	Pred.	No. 2, 6e-48;
Matches		146;	Conservative	70; Mismatches 146; Indels 34; Gaps 15;
QY	66	GLIGIRHDFC-SCKYRCKSSFCIETIARCDGSVSDCKDQGDEYRCVRY---GGONAVLQVF 120	RL	"Genomic organization of the human enterokinase.";
DB	652	GLGIPPEPKEDNFQCKDG-ECIPLVNLCDGFPHCKDGSDEAHCVRLFNGTDSSGLVQFR 710	RN	Submitted (UN-1999) to the EMBL/GenBank/DDBJ databases.
QY	121	TAASWMTMCSDIWKGHIANVACAQIQLGPSPVSSDNLURVSLEQGREFEFSIDHLIPDDK 180	RA	[3]
DB	711	IQSISWIVACAEWTQIISDQVYDQIGT--GNSSVPTFSTGG--GPYVNLN----- 758	RA	SEQUENCE FROM N.A.
QY	181	VTLAHIVSIV-REGASGHVTLQCT--AGCHR--RGYSSRIVGGMNSLISQWQAS 233	RA	RA
DB	759	-TAPNGSLLTPSQOCLEDSILLQCNQYKSGKKLYTQEVSPIKVGSDSREGARWVVA 817	RA	Rehmsmeier S., Borzym K., Gardiner K., Nizetic D., Francis F.,
QY	234	LOFGYHLLGGSVITPLWITTAHYCVYDLY-PKSMTIQVSL--VSLIDNP-APSHLVEK 289	RA	Shintani A., Sasaki T., Niedamine K., Mitsuyama S., Antonarakis S.E.,
DB	818	LYFDDQGQVCGASLVSRODWLVSAAHCVYGRNMPEPSKWKAVLGHMASNLTSPOIETRLIDQ 877	RA	Mitsushima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P., Blecker H.,
QY	290	IYHHSYKPKRKGNDIALMKGLAGPLTFNEMQPOVCLPNSSENFPODKYKWTSGWGATEDG 349	RA	Ranser J., Beck A., Klages S., Hennig S., Riessmann L., Dagand E.,
DB	878	IYVINPHYNKRKNNDIAMMILHEMVKVNTYDITOPICLPEENQVFPGRICSTAGWGAL-IV 936	RA	Weinmeier S., Reinhardt R., Yaspo M.-L.,
QY	350	AGDASPVNLHAPLNSLNKCNHR-DVYGGITSPSMCAG3LTGGVDSQCDSSGGPLVCC 408	RL	Lehrach H., Reinhardt R., Yaspo M.-L.,
DB	937	OGSTADVLQKADYPLSNEKQQMPEYN-ITENNAVCAAGSEAGGDSCQDGGPLMCQ 994	RC	"The DNA sequence of human chromosome 21.";
QY	409	ERRLWKLVLGATSGFGICCAEVNPKPGVYTRVNTSFLDWI 444	RA	Kitamoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.,
DB	995	ENNRWILLAGTVTSFGYOCALPNRPGVYARVPRFTEWI 1030	RA	Rehmsmeier S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RESULT	10		RT	"Enterokinase, the initiator of intestinal digestion, is a mosaic
ENTK_HUMAN			RT	protease composed of a distinctive assortment of domains. ";
STANDARD			RT	Proc. Natl. Acad. Sci. U.S.A. 91:7589-7592 (1994).
PRT	1019 AA.		RC	-I- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
			CC	PROTEOLYTIC PROENZYMES (TRYPSIN), CHYMOTRYPSIN AND CARBOXYPEPTIDASE
			CC	A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
			CC	TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
			CC	PROCARBOXYPEPTIDASES, AND PROELASTASES.
			CC	-I- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
			CC	TRYPsinogen.
			CC	-I- SUBUNIT: HETEROODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
			CC	MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
			CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
			CC	-I- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
			CC	-I- TIM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
			CC	CLEARED BY A TRYPSIN-LIKE PROTEASE.
			CC	-I- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
			CC	MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
			CC	-I- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
			CC	-I- SIMILARITY: CONTAINS 2 CUB DOMAINS.
			CC	-I- SIMILARITY: CONTAINS 1 SEA DOMAIN.
			CC	-I- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
			CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
			CC	TRYPsin FAMILY.
			CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
			CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
			CC	the European Bioinformatics Institute. There are no restrictions on its
			CC	use by non-profit institutions as long as its content is in no way
			CC	modified and this statement is not removed. Usage by and for commercial
			CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
			CC	[1]
			DR	EMBL; U09860; AAC50138 1; .
			DR	EMBL; Y19125; CAB65555.1; JOINED.
			DR	EMBL; Y19125; CAB65555.1; JOINED.
			DR	EMBL; Y19127; CAB65555.1; JOINED.
			DR	EMBL; Y19128; CAB65555.1; JOINED.
			DR	EMBL; Y19129; CAB65555.1; JOINED.
			DR	EMBL; Y19130; CAB65555.1; JOINED.
			DR	EMBL; Y19131; CAB65555.1; JOINED.
			DR	EMBL; Y19132; CAB65555.1; JOINED.
			DR	EMBL; Y19133; CAB65555.1; JOINED.
			DR	EMBL; Y19134; CAB65555.1; JOINED.
			DR	EMBL; Y19135; CAB65555.1; JOINED.
			DR	EMBL; Y19136; CAB65555.1; JOINED.
			DR	EMBL; Y19137; CAB65555.1; JOINED.
			DR	EMBL; Y19138; CAB65555.1; JOINED.
			DR	EMBL; Y19139; CAB65555.1; JOINED.
			DR	EMBL; Y19140; CAB65555.1; JOINED.
			DR	EMBL; Y19141; CAB65555.1; JOINED.
			DR	EMBL; Y19142; CAB65555.1; JOINED.
			DR	EMBL; Y19143; CAB65555.1; JOINED.
			DR	EMBL; Y19144; CAB65555.1; JOINED.
			DR	EMBL; Y19145; CAB65555.1; JOINED.
			DR	EMBL; Y19146; CAB65555.1; JOINED.
			DR	EMBL; Y19147; CAB65555.1; JOINED.
			DR	EMBL; Y19148; CAB65555.1; JOINED.
			DR	EMBL; Y19149; CAB65555.1; JOINED.
			DR	EMBL; Y19150; CAB65555.1; JOINED.
			DR	EMBL; Y19151; CAB65555.1; JOINED.
			DR	EMBL; Y19152; CAB65555.1; JOINED.
			DR	EMBL; Y19153; CAB65555.1; JOINED.
			DR	EMBL; Y19154; CAB65555.1; JOINED.
			DR	EMBL; Y19155; CAB65555.1; JOINED.
			DR	EMBL; Y19156; CAB65555.1; JOINED.
			DR	EMBL; Y19157; CAB65555.1; JOINED.
			DR	EMBL; Y19158; CAB65555.1; JOINED.
			DR	EMBL; Y19159; CAB65555.1; JOINED.
			DR	EMBL; Y19160; CAB65555.1; JOINED.
			DR	EMBL; Y19161; CAB65555.1; JOINED.
			DR	EMBL; Y19162; CAB65555.1; JOINED.
			DR	EMBL; Y19163; CAB65555.1; JOINED.
			DR	EMBL; Y19164; CAB65555.1; JOINED.
			DR	EMBL; Y19165; CAB65555.1; JOINED.
			DR	EMBL; Y19166; CAB65555.1; JOINED.
			DR	EMBL; Y19167; CAB65555.1; JOINED.
			DR	EMBL; Y19168; CAB65555.1; JOINED.
			DR	EMBL; Y19169; CAB65555.1; JOINED.
			DR	EMBL; Y19170; CAB65555.1; JOINED.
			DR	EMBL; Y19171; CAB65555.1; JOINED.
			DR	EMBL; Y19172; CAB65555.1; JOINED.
			DR	EMBL; Y19173; CAB65555.1; JOINED.
			DR	EMBL; Y19174; CAB65555.1; JOINED.
			DR	EMBL; Y19175; CAB65555.1; JOINED.
			DR	EMBL; Y19176; CAB65555.1; JOINED.
			DR	EMBL; Y19177; CAB65555.1; JOINED.
			DR	EMBL; Y19178; CAB65555.1; JOINED.
			DR	EMBL; Y19179; CAB65555.1; JOINED.
			DR	EMBL; Y19180; CAB65555.1; JOINED.
			DR	EMBL; Y19181; CAB65555.1; JOINED.
			DR	EMBL; Y19182; CAB65555.1; JOINED.
			DR	EMBL; Y19183; CAB65555.1; JOINED.
			DR	EMBL; Y19184; CAB65555.1; JOINED.
			DR	EMBL; Y19185; CAB65555.1; JOINED.
			DR	EMBL; Y19186; CAB65555.1; JOINED.
			DR	EMBL; Y19187; CAB65555.1; JOINED.
			DR	EMBL; Y19188; CAB65555.1; JOINED.
			DR	EMBL; Y19189; CAB65555.1; JOINED.
			DR	EMBL; Y19190; CAB65555.1; JOINED.
			DR	EMBL; Y19191; CAB65555.1; JOINED.
			DR	EMBL; Y19192; CAB65555.1; JOINED.
			DR	EMBL; Y19193; CAB65555.1; JOINED.
			DR	EMBL; Y19194; CAB65555.1; JOINED.
			DR	EMBL; Y19195; CAB65555.1; JOINED.
			DR	EMBL; Y19196; CAB65555.1; JOINED.
			DR	EMBL; Y19197; CAB65555.1; JOINED.
			DR	EMBL; Y19198; CAB65555.1; JOINED.
			DR	EMBL; Y19199; CAB65555.1; JOINED.
			DR	EMBL; Y19200; CAB65555.1; JOINED.
			DR	EMBL; Y19201; CAB65555.1; JOINED.
			DR	EMBL; Y19202; CAB65555.1; JOINED.
			DR	EMBL; Y19203; CAB65555.1; JOINED.
			DR	EMBL; Y19204; CAB65555.1; JOINED.
			DR	EMBL; Y19205; CAB65555.1; JOINED.
			DR	EMBL; Y19206; CAB65555.1; JOINED.
			DR	EMBL; Y19207; CAB65555.1; JOINED.
			DR	EMBL; Y19208; CAB65555.1; JOINED.
			DR	EMBL; Y19209; CAB65555.1; JOINED.
			DR	EMBL; Y19210; CAB65555.1; JOINED.
			DR	EMBL; Y19211; CAB65555.1; JOINED.
			DR	EMBL; Y19212; CAB65555.1; JOINED.
			DR	EMBL; Y19213; CAB65555.1; JOINED.
			DR	EMBL; Y19214; CAB65555.1; JOINED.
			DR	EMBL; Y19215; CAB65555.1; JOINED.
			DR	EMBL; Y19216; CAB65555.1; JOINED.
			DR	EMBL; Y19217; CAB65555.1; JOINED.
			DR	EMBL; Y19218; CAB65555.1; JOINED.
			DR	EMBL; Y19219; CAB65555.1; JOINED.
			DR	EMBL; Y19220; CAB65555.1; JOINED.
			DR	EMBL; Y19221; CAB65555.1; JOINED.
			DR	EMBL; Y19222; CAB65555.1; JOINED.
			DR	EMBL; Y19223; CAB65555.1; JOINED.
			DR	EMBL; Y19224; CAB65555.1; JOINED.
			DR	EMBL; Y19225; CAB65555.1; JOINED.
			DR	EMBL; Y19226; CAB65555.1; JOINED.
			DR	EMBL; Y19227; CAB65555.1; JOINED.
			DR	EMBL; Y19228; CAB65555.1; JOINED.
			DR	EMBL; Y19229; CAB65555.1; JOINED.
			DR	EMBL; Y19230; CAB65555.1; JOINED.
			DR	EMBL; Y19231; CAB65555.1; JOINED.
			DR	EMBL; Y19232; CAB65555.1; JOINED.
			DR	EMBL; Y19233; CAB65555.1; JOINED.
			DR	EMBL; Y19234; CAB65555.1; JOINED.
			DR	EMBL; Y19235; CAB65555.1; JOINED.
			DR	EMBL; Y19236; CAB65555.1; JOINED.
			DR	EMBL; Y19237; CAB65555.1; JOINED.
			DR	EMBL; Y19238; CAB65555.1; JOINED.
			DR	EMBL; Y19239; CAB65555.1; JOINED.
			DR	EMBL; Y19240; CAB65555.1; JOINED.
			DR	EMBL; Y19241; CAB65555.1; JOINED.
			DR	EMBL; Y19242; CAB65555.1; JOINED.
			DR	EMBL; Y19243; CAB65555.1; JOINED.
			DR	EMBL; Y19244; CAB65555.1; JOINED.
			DR	EMBL; Y19245; CAB65555.1; JOINED.
			DR	EMBL; Y19246; CAB65555.1; JOINED.
			DR	EMBL; Y19247; CAB65555.1; JOINED.
			DR	EMBL; Y19248; CAB65555.1; JOINED.
			DR	EMBL; Y19249; CAB65555.1; JOINED.
			DR	EMBL; Y19250; CAB65555.1; JOINED.
			DR	EMBL; Y19251; CAB65555.1; JOINED.
			DR	EMBL; Y19252; CAB65555.1; JOINED.
			DR	EMBL; Y19253; CAB65555.1; JOINED.
			DR	EMBL; Y19254; CAB65555.1; JOINED.
			DR	EMBL; Y19255; CAB65555.1; JOINED.
			DR	EMBL; Y19256; CAB65555.1; JOINED.
			DR	EMBL; Y19257; CAB65555.1; JOINED.
			DR	EMBL; Y19258; CAB65555.1; JOINED.
			DR	EMBL; Y19259; CAB65555.1; JOINED.
			DR	EMBL; Y19260; CAB65555.1; JOINED.
			DR	EMBL; Y19261; CAB65555.1; JOINED.
			DR	EMBL; Y19262; CAB65555.1; JOINED.
			DR	EMBL; Y19263; CAB65555.1; JOINED.
			DR	EMBL; Y19264; CAB65555.1; JOINED.
			DR	EMBL; Y19265; CAB65555.1; JOINED.
			DR	EMBL; Y19266; CAB65555.1; JOINED.
			DR	EMBL; Y19267; CAB65555.1; JOINED.
			DR	EMBL; Y19268; CAB65555.1; JOINED.
			DR	EMBL; Y19269; CAB65555.1; JOINED.
			DR	EMBL; Y19270; CAB65555.1; JOINED.
			DR	EMBL; Y19271; CAB65555.1; JOINED.
			DR	EMBL; Y19272; CAB65555.1; JOINED.
			DR	EMBL; Y19273; CAB65555.1; JOINED.
			DR	EMBL; Y19274; CAB65555.1; JOINED.
			DR	EMBL; Y19275; CAB65555.1; JOINED.
			DR	EMBL; Y19276; CAB65555.1; JOINED.
			DR	EMBL; Y19277; CAB65555.1; JOINED.
			DR	EMBL; Y19278; CAB65555.1; JOINED.
			DR	EMBL; Y19279; CAB65555.1; JOINED.
			DR	EMBL; Y19280; CAB65555.1; JOINED.
			DR	EMBL; Y19281; CAB65555.1; JOINED.
			DR	EMBL; Y19282; CAB65555.1; JOINED.
			DR	EMBL; Y19283; CAB65555.1; JOINED.
			DR	EMBL; Y19284; CAB65555.1; JOINED.
			DR	EMBL; Y19285; CAB65555.1; JOINED.
			DR	EMBL; Y19286; CAB65555.1; JOINED.
			DR	EMBL; Y19287; CAB65555.1; JOINED.
			DR	EMBL; Y19288; CAB65555.1; JOINED.
			DR	EMBL; Y19289; CAB65555.1; JOINED.
			DR	EMBL; Y19290; CAB65555.1; JOINED.
			DR	EMBL; Y19291; CAB65555.1; JOINED.
			DR	EMBL; Y19292; CAB65555.1; JOINED.
			DR	EMBL; Y19293; CAB65555.1; JOINED.
			DR	EMBL; Y19294; CAB65555.1; JOINED.
			DR	EMBL; Y19295; CAB65555.1; JOINED.
			DR	EMBL; Y19296; CAB65555.1; JOINED.
			DR	EMBL; Y19297; CAB65555.1; JOINED.
			DR	EMBL; Y19298; CAB65555.1; JOINED.
			DR	EMBL; Y19299; CAB65555.1; JOINED.
			DR	EMBL; Y19300; CAB65555.1; JOINED.
			DR	EMBL; Y19301; CAB65555.1; JOINED.
			DR	EMBL; Y19302; CAB65555.1;

RA Matsushima M., Ichinose M., Yagami N., Kakei N., Tsukada S.,
 RA Miki K., Kurokawa K., Tashiro K., Shiokawa K., Shinomiya K.,
 RA Uneyama H., Inoue H., Takahashi T., Takahashi K.;
 RT "Structural characterization of porcine enteropeptidase.";
 RL J. Biol. Chem. 269:19976-19982 (1994).
 CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC -!- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPSINOGEN.
 CC -!- SUBUNIT: HETERTRIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
 CC (HEAVY) CHAIN, AND A MINI CHAIN.
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
 CC -!- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
 CC -!- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use. By non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 EMBL: D30799; BA06459.1; -
 HSSP: P0763; 1DPC.
 DR M6RPS; S01.156; -
 DR InterPro; IPR00859; CUB.
 DR InterPro; IPR01314; Chymotrypsin.
 DR InterPro; IPR02172; LDL_recept_A.
 DR InterPro; IPR00998; MAM.
 DR InterPro; IPR00082; SEA.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001254; trypsin.
 DR Pfam; PF000431; CUB; 2.
 DR Pfam; PF00057; IgI_recept_a; 2.
 DR Pfam; PF00029; MAM; 1.
 DR Pfam; PF01390; SEA; 1.
 DR Pfam; PF00530; SRCR; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0722; CHYMOTRYPSIN.
 DR SMART; SM00042; CUB; 2.
 SMART; SM00192; LDDA; 2.
 SMART; SM00137; MAM; 1.
 SMART; SM00200; SEA; 1.
 SMART; SM00202; SR; 1.
 SMART; SM00200; TRY_SPC; 1.
 PROSITE; PS01180; CUB; 2.
 PROSITE; PS01209; LDIKA_1; 2.
 PROSITE; PS00068; LDIKA_2; 2.
 PROSITE; PS0070; MAM_1; 1.
 PROSITE; PS50060; MAM_2; 1.
 PROSITE; PS50024; SEA; 1.
 PROSITE; PS00042; SRCR_1; FALSE_NEG.
 PROSITE; PS50240; TRYPSIN_DOM; 1.
 PROSITE; PS00134; TRYPSIN_HIS; 1.
 PROSITE; PS500135; TRYPSIN_SER; 1.
 KW Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 Serine Protease; Zymogen; Transmembrane; Repeat.
 CHAIN 52 117 NON-CATALYTIC M CHAIN (MINI CHAIN).
 FT CHAIN 118 799 NON-CATALYTIC H CHAIN (HEAVY CHAIN).
 CHAIN 800 1034 CATALYTIC L CHAIN (LIGHT CHAIN).

FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 52 169 SEA.
 FT DOMAIN 52 169 LDI-RECEPTOR CLASS A 1.
 FT DOMAIN 197 238 CUB 1.
 FT DOMAIN 240 349
 FT DOMAIN 357 519
 FT DOMAIN 539 649 CUB 2.
 FT DOMAIN 656 694 LDI-RECEPTOR CLASS A 2.
 FT DOMAIN 693 786 SERC.
 FT DOMAIN 800 1034 SERINE PROTEASE.
 FT DOMAIN 840 840 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 891 891 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 986 986 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT LIPID 986 986 MYristate (POTENTIAL).
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 206 225 BY SIMILARITY.
 FT DISULFID 219 236 BY SIMILARITY.
 FT DISULFID 658 670 BY SIMILARITY.
 FT DISULFID 665 683 BY SIMILARITY.
 FT DISULFID 677 692 BY SIMILARITY.
 FT DISULFID 787 911 BY SIMILARITY.
 FT DISULFID 825 841 BY SIMILARITY.
 FT DISULFID 925 992 BY SIMILARITY.
 FT DISULFID 956 971 BY SIMILARITY.
 FT DISULFID 982 1010 BY SIMILARITY.
 FT CARBOHYD 116 116 BY SIMILARITY.
 FT CARBOHYD 147 147 BY SIMILARITY.
 FT CARBOHYD 170 170 BY SIMILARITY.
 FT CARBOHYD 194 194 BY SIMILARITY.
 FT CARBOHYD 283 283 BY SIMILARITY.
 FT CARBOHYD 343 343 BY SIMILARITY.
 FT CARBOHYD 350 350 BY SIMILARITY.
 FT CARBOHYD 403 403 BY SIMILARITY.
 FT CARBOHYD 455 455 BY SIMILARITY.
 FT CARBOHYD 485 485 BY SIMILARITY.
 FT CARBOHYD 518 518 BY SIMILARITY.
 FT CARBOHYD 549 549 BY SIMILARITY.
 FT CARBOHYD 645 645 BY SIMILARITY.
 FT CARBOHYD 679 697 BY SIMILARITY.
 FT CARBOHYD 701 701 BY SIMILARITY.
 FT CARBOHYD 722 721 BY SIMILARITY.
 FT CARBOHYD 740 740 BY SIMILARITY.
 FT CARBOHYD 761 761 BY SIMILARITY.
 FT CARBOHYD 804 804 BY SIMILARITY.
 FT CARBOHYD 863 863 BY SIMILARITY.
 FT CARBOHYD 902 902 BY SIMILARITY.
 FT CARBOHYD 964 964 BY SIMILARITY.
 FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1034 AA; 114776 MW; 038BC64CF64CC368 CRC64;

Query Match 26.7%; Score 654; DB 1; Length 1034;
 Best Local Similarity 37.7%; Pred. No. 8.5e-48; Gaps 15;
 Matches 149; Conservative 65; Mismatches 147; Indels 34;

Qy 67 LGIFHDFC-SGKRCRCSRFFKCTELIARDGVSCKDGDEYRVR---VGGSONAVIQLFV 121
 Db 652 LGIPEPKEDNFQCG-ECVILVNLCGDSFCKGDSDEAHCVRLFTGANNGLVQFRI 710
 Qy 122 AASWKTMCSSDMKGHAYANVACQLGFFPSVYSSDNRLVSSLCQFREEFVSDHLLPDDKV 181
 Db 711 QSIWHTACAEWNTTOPDDCQCOLLGLGIGTGNNSMPPF-FSSGG---P-FVKIN----- 757
 Qy 182 TALHHSVYV--REGCASGHVVLQCT--ACGGR--RGYSSRIVGGNMSLSQLWPQASL 234
 Db 758 TAPNGSLITASEQCFEDSLILQLCNHKSGKQVAEVSPKIVGGNDREGAWPWAVAL 817
 Qy 235 QPGYHICGGSVITPLMIIITAHCYVFLYL-PKSWTIQLVGL--VSLLDNP-APSHVEK 290
 Db 818 YNGQOLICGASLVRSDWLVSAACVVGGRNLSKWAIGLMTSNTSPQIVTRLDEI 877
 Qy 291 VHSKVKRKLGLDIALMLKLAGPLTNEIMQPCLNSENEPDGVKWCWSGATEDGA 350
 Db 878 VNPVHYRKRKSDIAMLMMHLFEPKVNTDYIOPICLDEENQFPPGRICSAGWGVY-IYQ 936

DR	InterPro; IPR001314; Chymotrypsin.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DR	InterPro; IPR001190; SRCR.	OC	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Rattus.
DR	InterPro; IPR001254; Trypsin.	OX	NCBI_TaxID=10116;
DR	pfam; PF00089; trypsin_1.	RN	[1]
DR	PRINTS; PR00722; CHYMO TRYPSIN.	RP	SEQUENCE FROM N.A.
DR	SMART; SM00022; SR_1.	RX	MEDLINE:91129236; PubMed=1993180;
DR	SMART; SM00020; TRYSPC_1.	RA	Beubien G., Rosinski-Chupin I., Mattei M.-G., Mbikay M., Chretien M.,
DR	PROSITE; PS00240; TRYPSIN_DOM_1.	RA	Seiddah N.G.;
DR	PROSITE; PS00134; TRYPSIN_HIS_1.	RA	Rougeon F., Lazare C., Chretien M.;
DR	PROSITE; PS00135; TRYPSIN_SER_1.	RT	"The cDNA structure of rat plasma kallikrein.;"
KW	Hydrolyase; Serine protease; Transmembrane; Signal-anchor.	RL	Biochemistry 30:1628-1635(1991).
FT	CHAIN	RN	[2]
FT	CHAIN	RX	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
FT	CHAIN	RA	MEDLINE=90091743; PubMed=2598771;
FT	DOMAIN	RA	Seiddah N.G., Ladenheim R., Mbikay M., Hamelin J., Lutfalla G.,
FT	TRANSMEM	RA	Rougeon F., Lazare C., Chretien M.;
FT	DOMAIN	RT	"The cDNA structure of rat plasma kallikrein.;"
FT	DOMAIN	RL	DNA 8:563-574(1989).
FT	ACT_SITE	CC	-1- FUNCTION: THE ENZYME CLEAVES LYS ARG AND ARG-SER BONDS. IT
FT	ACT_SITE	CC	ACTIVATES IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
FT	ACT_SITE	CC	TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRAKININ FROM
FT	DISULFID	CC	HMW KINTININ AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
FT	DISULFID	CC	SYSTEM BY CONVERTING PRORENIN INTO RENIN.
FT	DISULFID	CC	-1- SUBUNIT: THE ZIMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
FT	DISULFID	CC	THE MOLECULE INTO A LIGHT CHAIN, WHICH ASSOCIATES WITH HMW KINTININ. THESE
FT	DISULFID	CC	AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KINTININ. THESE
FT	CARBONYL	CC	CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
FT	SEQUENCE	CC	-1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
FT	SEQUENCE	CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
FT	SEQUENCE	CC	TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
FT	SEQUENCE	CC	-----
FT	SEQUENCE	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
FT	SEQUENCE	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
FT	SEQUENCE	CC	the European Bioinformatics Institute. There are no restrictions on its
FT	SEQUENCE	CC	use by non-profit institutions as long as its content is in no way
FT	SEQUENCE	CC	modified and this statement is not removed. Usage by and for commercial
FT	SEQUENCE	CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
FT	SEQUENCE	CC	-----
QY	60 TLLALAGIGIHFDCSGKRYCRCSSEFKCIELIARDGDYVSDCKDGEDEYRCVRYGGNAVLOV 119	DR	EMBL; M62357; AAA74563.1; JOINED.
Db	66 DKTEGTWTWLICSSRSNARVAGLGCEENGFLRALLAHSELDDVRTAGANTGTSFFCVDDEGLP 125	DR	EMBL; M62358; AAA74563.1; JOINED.
Db	17 VAWLWVOTLFLFGIG--AASWAIVILLR----SDOEPVQVQLSPGDSRLVL 65	DR	EMBL; M62346; AAA74563.1; JOINED.
Qy	120 -FTAAASKTMCSDDWKGHYANVACAOGLGPSPVYSSDNLRLVSSLEGDFREEVYSD---- 173	DR	EMBL; M62347; AAA74563.1; JOINED.
Db	126 LAQRLL--DVTISWC-----DCPGRGRFLPATCOPCGRRKLPWDRIVGGQDSSLGRWPW 230	DR	EMBL; M62349; AAA74563.1; JOINED.
Db	231 QASLQFGSYHLCGGSVITPLWITTAIAHVYDLYLP-----SWTLOVGLVSLLDNPAPS 284	DR	EMBL; M62350; AAA74563.1; JOINED.
Db	176 QVSLRYGTHLCCGSSLSDGWLTAAHC---FPERNRVLRSWRWRVAGAVARTSHAVO 230	DR	EMBL; M62351; AAA74563.1; JOINED.
Qy	285 HLVEKTVYHVKYKPR-----LGNDITALMKIAGPLTFNEMIOPVCLPNSEBNFPGKVC 338	DR	EMBL; M62352; AAA74563.1; JOINED.
Db	231 LGVQAVIYHGGVLPFRDPTIDENSNDIALVHLVSSLPLTEVQOPVCLPAAGALVDGKVC 290	DR	EMBL; M62353; AAA74563.1; JOINED.
Qy	339 WISGKGATEDGAGDASVILNHAAPLTSNPKCNHRVYGGIISPSMCAGLIGGGVDSQ 398	DR	EMBL; M62354; AAA74563.1; JOINED.
Db	291 TWTGWNHQ--FYGQAVVLPQEARVPTISNEVYCNNSPDPYGNQIKPKMFCAQYPEGIDACQ 349	DR	EMBL; M62355; AAA74563.1; JOINED.
Qy	399 GDSGGPLVQCR--RLWKLVGATSTFGIGGAENPKGVYTRVTSFLDWHQME 449	DR	EMBL; M62356; AAA74563.1; JOINED.
Db	350 GDSGGHRYCVEDRISGTSRWRUCCGIVSWGTCALARPGVYITKVIDREWIQAQK 404	DR	EMBL; M62357; AAA74563.1; JOINED.
RESULT	14	DR	M3002; AAA4163.1; -.
KAL_RAT	KAL_RAT	DR	M58590; AAA42069.1; -.
ID	STANDARD	DR	PTR; A39180; KORIPL.
AC	P14272;	DR	HSSP; P00750; KRF.
DT	01-JAN-1990 (Rel. 13, Created)	DR	MEOPS; S01.212; -.
DT	01-JAN-1990 (Rel. 13, Last sequence update)	DR	INTERPRO; IPR00177; Apple.
DT	20-AUG-2001 (Rel. 40, Last annotation update)	DR	INTERPRO; IPR01314; Chymotrypsin.
DE	PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)	DR	INTERPRO; IPR003014; PAN.
DE	(KININOGENTIN) (FLETCHER FACTOR)	DR	INTERPRO; IPR001254; Trypsin.
GN	PK	DR	INTERPRO; IPR00024; PAN; 4
OS	Rattus norvegicus (Rat).	DR	INTERPRO; IPR0009; Trypsin; 1.
		DR	PRINTS; PR00005; APPLE DOMAIN.
		DR	PRINTS; PR00722; CHYMO TRYPSIN.
		DR	SMART; SM0023; Apple; 4.
		DR	SMART; SM00495; Apple; 4.
		DR	PROSITE; PS00049; TRYSPC_1.
		DR	PROSITE; PS00240; TRYPSIN_DOM_1.
		DR	INTERPRO; IPR00134; Trypsin HIS; 1.
		DR	PROSITE; PS00135; TRYPSIN SER; 1.
		DR	Hydrolyase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
		DR	Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
		DR	Repeat.
		FT	SIGNAL; 1 19

FT	DISULFID	111	194	BY SIMILARITY.
FT	DISULFID	137	166	BY SIMILARITY.
FT	DISULFID	141	147	BY SIMILARITY.
FT	DISULFID	201	284	BY SIMILARITY.
FT	DISULFID	227	256	BY SIMILARITY.
FT	DISULFID	231	237	BY SIMILARITY.
FT	DISULFID	292	375	BY SIMILARITY.
FT	DISULFID	318	347	BY SIMILARITY.
FT	DISULFID	322	328	BY SIMILARITY.
FT	DISULFID	340	345	BY SIMILARITY.
FT	DISULFID	383	503	BY SIMILARITY.
FT	DISULFID	419	435	BY SIMILARITY.
FT	DISULFID	517	584	BY SIMILARITY.
FT	DISULFID	548	563	BY SIMILARITY.
FT	DISULFID	574	602	BY SIMILARITY.
SQ	SEQUENCE	638	AA:	671368 MW: CC27C939AB1086599 CRC64;
Query Match				
Best	Local Similarity	22.9%	Score	561;
Matches	126;	Conservative	34.8%	Pred. No. 4e-40;
			61;	Mismatches 109;
				Indels 66;
				Gaps 12;
QY	DRKGHYANWACAOGLGFPSVYVSSDNRLRVSSLESQFREEFVSDIHDLLPDDKRTVTAHLHSVVV	191		
Db	301 DEGEELENV-----TVOGADVCOETCTKTIRCPFF-TYSLIPQD-----CNE	342		
QY	192 EGC-----ASGHWVTL-----QCTACGHRRGGYSSRIVGGNN	223		
Db	343 BGCKCSLRLSTBDGSPTRITYGMQGSSGYSLRICKLVLDSPDCTT-----KINARIVGGNN	397		
QY	224 LLSQWPWQASLQFQ--GYHLCGGSVITPLWITTAHCVYDLYLPKSWTIQVGLSVL--L	278		
Db	398 SIGEWMPQVSLQVLYKLSQTHLOGGSIIGRQWVITAHCEDGIRPYDWRVYIGGLSLSI	457		
QY	279 DNAPASHVTEKIVYHSKVKPKRLGLNDLAMKLGQPLTNEMTQPVCLPNSEENPDGKVC	338		
Db	458 TKEPTSSRKEKELIHOEKVKSEGNYDIALKLQPLNTEFQPKICLUSSKADNTIYTC	517		
QY	339 WTSWGWTATEDGAGDASPVLNHAAPLISNKIC--NHRDVGIGIISPSCMCAGYLGGVDS	396		
Db	518 WWTGWGTYKE--QGETQONLILQKATIPLVNEECOKYRDY--VINKQMCAGYKEGTDIA	573		
QY	397 CGDGSGLPVLQCBERRLWLLVGATSGFGICCAEWKPGVITRVSFLDWHHEOME-- RDL	452		
Db	574 CKGDGSPPLVCKHSGRQWLVGITSWGECCGRKDQPGVYTKVSEYMDWLEKTIQSSDVRAL	633		
QY	453 KT	454		
Db	634 ET	635		

Search completed: April 11, 2002, 08:57:04
Job time: 217 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:52:32 ; Search time 25.01 Seconds
(without alignments)

1344.633 Million cell updates/sec

Title: US-09-846-512-12
perfect score: 2447
Sequence: I MGENDPPAVEAPFSSFRSLFG. TRVTSFLDWTHQMERDLKT 4 54
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%, Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDS8/gcgatata/geneseq/geneseq/AA1980.DAT: *
2: /SIDS8/gcgatata/geneseq/geneseq/AA1981.DAT: *
3: /SIDS8/gcgatata/geneseq/geneseq/AA1982.DAT: *
4: /SIDS8/gcgatata/geneseq/geneseq/AA1983.DAT: *
5: /SIDS8/gcgatata/geneseq/geneseq/AA1984.DAT: *
6: /SIDS8/gcgatata/geneseq/AA1985.DAT: *
7: /SIDS8/gcgatata/geneseq/AA1986.DAT: *
8: /SIDS8/gcgatata/geneseq/AA1987.DAT: *
9: /SIDS8/gcgatata/geneseq/AA1988.DAT: *
10: /SIDS8/gcgatata/geneseq/AA1989.DAT: *
11: /SIDS8/gcgatata/geneseq/AA1990.DAT: *
12: /SIDS8/gcgatata/geneseq/AA1991.DAT: *
13: /SIDS8/gcgatata/geneseq/AA1992.DAT: *
14: /SIDS8/gcgatata/geneseq/AA1993.DAT: *
15: /SIDS8/gcgatata/geneseq/AA1994.DAT: *
16: /SIDS8/gcgatata/geneseq/AA1995.DAT: *
17: /SIDS8/gcgatata/geneseq/AA1996.DAT: *
18: /SIDS8/gcgatata/geneseq/AA1997.DAT: *
19: /SIDS8/gcgatata/geneseq/AA1998.DAT: *
20: /SIDS8/gcgatata/geneseq/AA1999.DAT: *
21: /SIDS8/gcgatata/geneseq/AA2000.DAT: *
22: /SIDS8/gcgatata/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES		OS		ALIGNMENTS	
Result No.	Score	Query Match Length	DB ID	Description	
1	2432.5	99.4	453	22 AAE06935	Human membrane-type
2	240.5	99.3	453	21 AAB44250	Human PRO382 (UNQ3
3	241.8	98.8	452	20 AAY41694	Human PRO382 prote
4	237.5	97.1	454	21 AAB32246	Tumour associated
5	178.8	73.1	327	21 AAY7203	Human serine prote
6	135.4	55.3	294	21 AAB32247	Tumour associated
7	133.9	54.7	248	21 AAB43572	Human cancer assoc
8	103.4	42.3	193	22 AAG73966	Human colon cancer
9	87.6	35.8	492	21 AAY92050	HRPc6/7 polypepti
10	87.6	35.8	492	22 AAM01315	Human 1000C amino acid
11	87.5	35.8	492	21 AAY77726	Human tumour suppr

12	87.5	35.8	492	21 AAY44406	Human 20PPI12-GRNC2
13	868.5	35.5	393	22 AAM01317	Human 1000C partial ami
14	864.5	35.3	492	21 AAB36901	Human ThPRSS2 prot
15	864.5	35.3	492	21 AAY57280	Ovrl15 homolog pro
16	864.5	35.3	492	22 AAB06943	Human serine prote
17	864.5	35.3	492	22 AAM01294	Human transmembran
18	73.1	29.9	421	22 AAB85042	Human SER6 protein
19	73.1	29.9	562	22 AAB01943	Human transmembran
20	72.7	29.7	688	22 AAB01944	Human transmembran
21	70.4	28.8	445	21 AAB11698	Mouse serine prote
22	70.4	28.8	480	21 AAB08950	Human secreted pro
23	70.2	28.7	414	21 AAB08912	Human secreted pro
24	69.6	28.4	457	21 AAB11699	Human serine prote
25	68.5	28.1	283	21 AAY81492	Human prostate-ass.
26	66.7	27.3	423	22 AAE06944	Human transmembran
27	66.4	27.2	417	22 AAE06942	Human hepsin prote
28	66.1	27.0	273	21 AAB11697	Mouse serine prote
29	66.0	27.0	435	22 AAY72558	Human protease HUP
30	66.0	27.0	31	15 AAY57283	Bovine enterokinase
31	66.0	27.0	798	15 AAY72559	Human serine prote
32	65.7	26.9	492	22 AAB6166	Human enterokinase
33	65.7	26.8	1019	22 AAE06940	Human membrane-typ
34	65.6	26.8	311	21 AAB11697	Mouse hepsin prote
35	65.5	26.8	416	20 AAY43325	Mouse hepsin prote
36	65.5	26.8	416	20 AYK96812	A mouse serine pro
37	65.2	26.6	432	21 AAY9417	Human PRO570 (UNQ
38	65.2	26.6	432	22 AAB87581	Human PRO570.
39	65.2	26.6	432	22 AAB6166	Protein of the inv
40	64.8	26.5	437	22 AAB06931	Human membrane-typ
41	64.6	26.4	172	22 AAM25245	Mouse serine prote
42	60.5	24.5	238	21 AAB11695	Human serine prote
43	583.5	23.8	317	22 AAM25633	Human protease
44	576	23.5	256	22 AAB98501	Human hepsin sequ
45	564.5	23.1	418	17 AAB89435	Trypsin-like enzym

XX	09-AUG-2001.	QY	421 FGIGCAEVNKGPGVYTRVTSFLDWIHEQMERDLKT 454
PD		Db	420 fgiqcaevnkgpgvtrvtfsfldwihqmerdlkt 453
XX	02-FEB-2001; 2001WO-US03471.		
XX	03-FEB-2000; 20000S-0179982.	RESULT	2
PR	18-FEB-2000; 20000S-0183542.	AABA4250	
PR	22-JUN-2000; 20000S-0213124.	ID	AABA4250 standard; Protein; 453 AA.
PR	20-JUL-2000; 20000S-0220970.	XX	
PR	03-SEP-2000; 20000S-0657986.	AC	AABA4250;
PR	22-SEP-2000; 20000S-0234840.	XX	
XX		DT	08-FEB-2001 (first entry)
PA	(CORY-) CORVAS INT INC.	XX	
XX		DE	Human PRO382 (UNQ323) Protein sequence SEQ ID NO:69.
PT	Madison EL, Ong EO, Yeh J;	XX	
XX	WPI; 2001-488877/53.	KW	Human; secreted protein; transmembrane protein; PRO; EST; cytostatic; expressed sequence tag; detection; cancer.
DR	N-PSDB; AAD13118.	XX	
XX		OS	Homo sapiens.
PT	Novel single chain polypeptide comprising protease domain of type-II membrane-type serine protease or its catalytically active portion useful for treating and preventing cancer and tumor	XX	
PT	XX	PN	WO20053756-A2.
PS	Claim 128; Page 214-215; 256pp; English.	XX	
XX		PD	14-SEP-2000.
CC	The invention relates to transmembrane serine proteases and their corresponding nucleotides and the protease domain of a type-II membrane-type serine protease (MTSP). MTSP is useful for identifying compounds that modulate or inhibits its proteolytic activity and for formulating a medicament for treating neoplastic disease. MTSP and its corresponding nucleotides are useful in preventing or treating tumours or cancers such as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in diagnostics and in hybridisation assays. MTSP is useful as a diagnostic marker for tumour development, growth and/or progression and as immunogens to generate antibodies that specifically bind to it. MTSP DNA is useful in a yeast two-hybrid system and in gene therapy. The present sequence is human MTSP6 protein.	PF	18-FEB-2000; 20000WO-US04341.
CC	XX	XX	
CC	Best Local Similarity 99.8%; pred. No. 4.1e-187; Matches 453; Conservative 0; Mismatches 0; Indels 1; Gaps 1;	PR	08-MAR-1999; 99WO-US05028.
QY	1 MGENDPAPVEAPPFSFRSLGDDQKISVAPDAAQOILSLPLKFPFTIVIGITALI 60	PR	12-MAR-1999; 99US-0123957.
Db	1 meendpapveapsfrsflgdklkisplkfpivigiali 60	PR	29-MAR-1999; 99US-0126773.
QY	61 LAUAGIGLIGHFDCSGKVRCSRKCIELIARQDGVSQCKDGEDYRCVRGGONAVLOVE 120	PR	21-APR-1999; 99US-0130232.
Db	61 laaigigihfdsgkycrccsfcieiarcdgvsackdgedeyrcvrvgggnaqvft 120	PR	28-APR-1999; 99US-0131445.
QY	121 TAAWKWICSDWIKGHYANVACAQIGPSVYSSDNLAVSLEQQFREEFVSDHLLPDK 180	PR	14-MAY-1999; 99US-0134287.
Db	121 taawsktmcddwkgbyanvacqigpsvssdnirvsslegqfrefvsidhlpdk 180	PR	23-JUN-1999; 99US-0141037.
QY	181 VRAUHHSVYVRECCASGHAVWVQCTAGCHRGYSSRIVGGNNMSSLSPWQASLQFOGH 240	PR	26-JUL-1999; 99US-0145698.
Db	181 vrahhsvyregcasaqhvrlqtacqaghrgqssrivgqansllsqpwqaslfqghy 240	PR	29-OCT-1999; 99US-0162506.
QY	241 ICGGSVITPLWITTAHCVYDYLPLPKSWTQVGLVSLIDNPAPSHLVEKTVHHSKYKPR 300	PR	30-NOV-1999; 99WO-US28313.
Db	241 logsvitplwitaahcvydlplkswtqvgivslldnppshlvekvhskypkr 300	PR	02-DEC-1999; 99WO-US28551.
Qy	301 IGDNDIAIMKLQPLFTHEMIQVCLPNEEENPDPKCYWTSGWATEDGAGDASPVLNA 360	PR	02-DEC-1999; 99WO-US28565.
Db	301 lgndialmkqlqplfthemiqvcplseenpdgkycwtsgwatedg-gdaspvlnha 360	PR	16-DEC-1999; 99WO-US00095.
QY	301 lgndialmkqlqplfthemiqvcplseenpdgkycwtsgwatedg-gdaspvlnha 359	PR	30-DEC-1999; 99WO-US31243.
Db	361 AVFLISKICNRDVGIGLISPSMCAGYLPGVGDSCQGDGSLVQERRWKLVGATS 420	PR	05-JAN-2000; 20000WO-US00219.
Db	360 avpliskicnrdvyygqispsmcagylpgvdsqsgpqlvqerrwkvgats 419	PR	06-JAN-2000; 20000WO-US00376.
PS	Sequence 453 AA.	XX	
XX		PA	(GERTH) GENENTECH INC.
XX		PI	Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
XX		PI	Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
XX		PI	Goldfarb A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillian KJ;
XX		PI	Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
XX		PI	Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX		DR	WPI; 2000-611443/58.
XX		DR	N-PSDB; AAC78475.
XX		PT	Novel PRO polypeptides and polynucleotides used in detection methods, to target bioactive molecules to specific cells, and to modulate cellular activities
XX		PS	Claim 12; Fig 28; 636pp; English.
XX		XX	
CC	ACT78458 to AAC78599 represent poly nucleotide and EST (expressed sequence tag) sequences which encode secreted or transmembrane PRO polypeptides. The PRO polynucleotides and polypeptides have cytostatic activity. The poly nucleotides and polypeptides can be used for detecting the presence of PRO polypeptides in samples, for linking bioactive molecules to cells and for modulating biological activities of cells, using the polypeptides for specific targeting. The polypeptide targeting can be used to kill the target cells, e.g. for the treatment of cancers. The polypeptide pairs provide specific targeting of bioactive molecules to cells. AAC78600 to AAC7897 represent PCR primers and probes used in the isolation of the PRO polynucleotide sequences.	CC	

SQ	Sequence	453 AA:
PR	20-MAR-1998;	98US-0078939
PR	25-MAR-1998;	98US-0079294
PR	26-MAR-1998;	98US-0079653
PR	27-MAR-1998;	98US-0079664
PR	27-MAR-1998;	98US-0079689
PR	27-MAR-1998;	98US-0079728
PR	27-MAR-1998;	98US-0079786
PR	27-MAR-1998;	98US-0079920
PR	30-MAR-1998;	98US-0079923
PR	31-MAR-1998;	98US-0080105
PR	31-MAR-1998;	98US-0080165
PR	31-MAR-1998;	98US-0080194
PR	01-APR-1998;	98US-0080328
PR	01-APR-1998;	98US-0080333
PR	08-APR-1998;	98US-0081049
PR	08-APR-1998;	98US-0081070
PR	09-APR-1998;	98US-0081195
PR	09-APR-1998;	98US-0081203
PR	15-APR-1998;	98US-0081817
PR	15-APR-1998;	98US-0081838
PR	15-APR-1998;	98US-0081955
PR	21-APR-1998;	98US-0082568
PR	22-APR-1998;	98US-0082700
PR	22-APR-1998;	98US-0082704
PR	22-APR-1998;	98US-0082767
PR	23-APR-1998;	98US-0082796
PR	23-APR-1998;	98US-0082796
PR	27-APR-1998;	98US-0083334
PR	28-APR-1998;	98US-0083322
PR	29-APR-1998;	98US-0083392
PR	29-APR-1998;	98US-0083496
PR	29-APR-1998;	98US-0083499
PR	29-APR-1998;	98US-0083545
PR	29-APR-1998;	98US-0083554
PR	29-APR-1998;	98US-0083558
PR	29-APR-1998;	98US-0083559
PR	30-APR-1998;	98US-0083742
PR	05-MAY-1998;	98US-0084199
PR	06-MAY-1998;	98US-0084414
PR	06-MAY-1998;	98US-0084441
PR	07-MAY-1998;	98US-0084554
PR	07-MAY-1998;	98US-0084600
PR	07-MAY-1998;	98US-0084627
PR	07-MAY-1998;	98US-0084637
PR	07-MAY-1998;	98US-0084639
PR	07-MAY-1998;	98US-0084640
PR	07-MAY-1998;	98US-0084643
PR	13-MAY-1998;	98US-0085233
PR	13-MAY-1998;	98US-0085338
PR	13-MAY-1998;	98US-0085339
PR	15-MAY-1998;	98US-0085573
PR	15-MAY-1998;	98US-0085700
PR	15-MAY-1998;	98US-0085704
PR	15-MAY-1998;	98US-0085580
PR	15-MAY-1998;	98US-0085582
PR	15-MAY-1998;	98US-0085689
PR	15-MAY-1998;	98US-0085697
PR	15-MAY-1998;	98US-0085700
PR	15-MAY-1998;	98US-0085700
PR	18-MAY-1998;	98US-0086023
PR	22-MAY-1998;	98US-0086392
PR	22-MAY-1998;	98US-0086414
PR	22-MAY-1998;	98US-0086430
PR	22-MAY-1998;	98US-0086446
PR	22-MAY-1998;	98US-0086446
PR	22-MAR-1998;	98US-0078936
PR	11-MAR-1998;	98US-0077632
PR	11-MAR-1998;	98US-0077641
PR	11-MAR-1998;	98US-0077649
PR	12-MAR-1998;	98US-0077791
PR	13-MAR-1998;	98US-0077004
PR	17-MAR-1998;	98US-004220
PR	20-MAR-1998;	98US-0078886
PR	20-MAR-1998;	98US-0078910
PR	20-MAR-1998;	98US-0078936
PR	121	TAASWKTMCDDWKGHTANVACAOQLGPRSYVSSDLNRVSLRGQFRBFRVSIDHLBDDK 180
PR	61	LALAIGLIGHIFCGSKGRCRSSFKCIELARCDGVSPCKDGDGEDEYRVVRGGONAVIQLW 120
PR	61	lalaiglighifcgskgrcrssfkcielarcdgvspckdgdgeleyrvvrqgavliqlw 120
PR	QY	YPLAHHSVYVREGCASHWVTLQCTACGHRRGYSSRIVGNNMSLSQLPQWQASLQFQGYH 240
PR	181	vtalhhsvyvregcashwvtlqctacghrgyssrviggnmslsgwpgas1qfgyh 240
PR	241	LCCGSVITPLWITAACWYDLYLPKSWTIQVGLVSUDNPAPSHLYEKIVVHSKPKR 300
PR	241	lgndialmkllagpltfnemipvcplnseendpgkrcwtsgwateqg-gdaspvlnha 359
PR	QY	AVPLISNCKINHRDVGGIIPSMLCAGYLGPGVDSQCGPLVYCOERPRAMKLGAT 420
PR	360	avplisknchrdvyyggiipsmlcagytggvdscqgdsqgplvcegerrlwlkvats 419
PR	QY	421 FGIGCAEWNPKGYVTRITSFLDWMIEQMERDIKT 454
PR	Db	420 fgigcaevnkpqyytrvtsfldwheqmerdikt 453
RESULT	3	
ID	AAY41694	
ID	AAY41694 standard; Protein; 452 AA.	
XX		
AC	AAY41694;	
XX		
DT	07-DEC-1999 (first entry)	
XX		
DE	Human PRO382 protein sequence.	
XX		
KW	Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation; probe; blood coagulation disorder; cancer; cellular adhesion disorder; secreted protein; transmembrane protein.	
XX		
OS	Homo sapiens.	
XX		
PN	W09946281-A2.	
XX		
PD	16-SEP-1999.	
XX		
PF	08-MAR-1999;	99WO-US05028.
XX		
PR	10-MAR-1998;	98US-0077450.
PR	11-MAR-1998;	98US-0077632.
PR	11-MAR-1998;	98US-0077641.
PR	11-MAR-1998;	98US-0077649.
PR	12-MAR-1998;	98US-0077791.
PR	13-MAR-1998;	98US-0077004.
PR	17-MAR-1998;	98US-004220
PR	20-MAR-1998;	98US-0078886.
PR	20-MAR-1998;	98US-0078910.
PR	20-MAR-1998;	98US-0078936.

PR	28-MAY-1998;	98US-0087098	ID	AAB32246 standard; Protein; 454 AA.
PR	28-MAY-1998;	98US-0087106	XX	
PR	28-MAY-1998;	98US-0087208	AC	AAB32246;
PR	30-JUL-1998;	98US-0094651	XX	
PR	11-SEP-1998;	98US-0100038	DT	11-JAN-2001 (first entry)
PA	(GBTH) GENENTECH INC.	XX	DE	Tumour associated differentially-expressed gene 12 protein sequence.
XX		XX	XX	Transmembrane serine protease; TAGD-12; chromosome 17; vaccination; tumour associated differentially-expressed gene 12; cytostatic; human; malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
PI	Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;	XX	KW	
XX	WPI: 1999-551358/46.	PN	KW	
DR	N-PSDB; AAZ333949.	XX	KW	
PS	New secreted and transmembrane polypeptides and their polynucleotides, useful for treating blood coagulation disorders, cancers and cellular adhesion disorders -	XX	OS	Homo sapiens.
PT	PT	XX	OS	
PT	New secreted and transmembrane polypeptides and their polynucleotides, useful for treating blood coagulation disorders, cancers and cellular adhesion disorders -	XX	OS	
PS	Claim 12; Fig 28; 530pp; English.	XX	OS	
XX		XX	OS	
CC	The present invention describes secreted and transmembrane polypeptides and their polynucleotides. The nucleotide sequences are useful as sources of probes, primers, for chromosome mapping, and for generation of antisense sequences. They can also be used to create transgenic animals. The proteins can be used to treat a variety of diseases and disorders, depending on their function. Diseases that may be treated include blood coagulation disorders, cancers and cellular adhesion disorders. They may also be used to raise antibodies. AAB33891 to AAZ34338, and AAY4174 represent polynucleotide and polypeptide sequence given in the exemplification of the present invention.	XX	OS	
CC	XX	XX	OS	
CC	Sequence 452 AA:	XX	OS	
Query Match	98.8%	Score 2418; DB 20; Length 452;	PS	02-MAR-2000; 2000WO-US5612.
Best Local Similarity	99.6%	Pred. No. 6e 186; 0; Mismatches 0; Indels 2; Gaps 2;	XX	03-MAR-1999; 99US-0261416.
Matches	452;	Conservative	XX	(UYAR-) UNIV ARKANSAS.
QY	1 MGENDPPAVEAPPSFRSLPGLDDLKISPVAPDAVAQIQLSLPLKFF-PIIVIGIALLI 60	CC	XX	O'Brien TJ, Underwood LJ;
Db	1 mgendppaveapsfrslkffdd-kispvapadavaaqiqlslplkffpiivigalll 59	CC	DR	WPI: 2000-533263/48.
QY	61 LALAIIGIHFEDSGKRYCRSFRCEIARCGDVSCKDGENEYRCVVGQNAQVF 120	CC	DR	N-PSDB; AAA93842.
Db	60 llaaigighfcsqsgkyrcrccsfkcielarcdgvsackdgedeyrcvrgqnavlq 119	CC	XX	DNA fragment encoding tumor associated differentially-expressed gene 12 protein used for diagnosing and treating malignant hyperplasia and cancers including ovarian cancer -
QY	121 TAAASKTMCSDDWKGHYANVACQOLGPSPSYVSSDLRVSLSLEGQFRERFVSDHLLDDK 180	CC	XX	PT
Db	120 taaskwtkmsdwwkgnyanvacqigfpsyssdnllivsslegqfreerfsidhlpdk 179	CC	XX	PT
QY	181 VTAALHHSVYVRECGCASHWVTOCTAGHRRYSSRIVGGNNMSLSQLPQWQASLQFOGYH 240	CC	XX	PS
Db	180 vtalhhsvyvregcaghvvtiqctacgrrhrgyssringqmsllsqwpwgastlqfgyh 239	CC	XX	Claim 3; Figure 4; 118pp; English.
QY	241 LCGGSVITPLWITIAKQVWDLKLPKSTIQLGVSLNDAAPSHLKEVINKISKPKR 300	CC	XX	This invention relates to a novel transmembrane serine protease called tumour associated differentially-expressed gene 12 (TAGD-12). TAGD-12 is located on chromosome 17. Sequences AAB93842-A93853 and AAB32246-B32249 represent human TAGD-12 cDNA and their corresponding protein sequences. A splice variant of TAGD-12 (TAGD-12V) leads to a truncated protein product. TAGD-12 is overexpressed in ovarian carcinomas. TAGD-12 exhibits cytostatic activity, and can be used in vaccines and in gene therapy. TAGD-12 nucleotide and protein sequences are used in the diagnosis of malignant hyperplasia and cancers or the ovary, breast, lung, colon, prostate and other cancers where TAGD-12 is overexpressed. TAGD-12 is particularly used as tumour marker for early disease diagnosis. TAGD-12 proteins or fragments can be used to vaccinate an individual with cancer, suspected of having a cancer or at risk of getting cancer. Sequences AAB93846-A93853 represent PCR primers used for amplifying the TAGD-12 cDNA sequence, and in the quantitative analysis of TAGD-12 mRNA. AAB32250 represents a peptide fragment of TAGD-12, used to create anti-TAGD-12 antibodies. Sequences AAB3251-32369 represent CC
Db	240 lcgssvtpwltitaanccvdylylpkswtqgqlslldpaphlivelvkyhskypkr 299	CC	XX	CC
QY	301 LQNDIALMKLAGLTENEMIOPVCLPNSBENFDGKVWTSGWATEGAGDASPVLNHA 360	CC	XX	CC
Db	300 lgndiamklagltfneimiqpcvlpsseenfdgkwtsgwatedg-gdaspvlnha 358	CC	XX	CC
QY	361 AVPLISIKICNRHWDYGGIISPSMMLCAGYLGGWDSGCGDSSGPPVCOERRLWKLKGATS 420	CC	XX	CC
Db	359 avplisikicnhdvyyggilspmsicagyltgvdscqgdsqgplvcoerrlwklkgats 418	CC	XX	CC
QY	421 FGCGCAEWINKPYWTRTSFEDWHEOMERDIT 454	CC	XX	CC
Db	419 fgigcaevnkpqpytrvtsflowhemedikt 452	CC	XX	CC
RESULT	4			
AB32246				
Query Match	97.1%	Score 2375; DB 21; Length 454;	QY	60 LALAIIGIHFEDSGKRYCRSFRCEIARCGDVSCKDGENEYRCVVGQNAQVF 119
Best Local Similarity	97.6%	Pred. No. 1.7e-182; 2; Mismatches 7; Indels 2; Gaps 2;	Db	61 llaaigighfcsqsgkyrcrccsfkcieltrdgvsackdgedeyrcvrgqnavlq 120
Matches	444;	Conservative	QY	120 TAAASKTMCSDDWKGHYANVACQOLGPSPSYVSSDLRVSLSLEGQFRERFVSDHLLDD 179
QY	1 MGENDPPAVEAPPSFRSLPGLDDLKISPVAPDAVAQIQLSLPLKFF-PIIVIGIALLI 59	Db	121 taaskwtkmsdwwkgnyanvacqigfpsyssdnllivsslegqfreerfsidhlpd 180	

```

QY 180 KVTALHISVYVERGAGSGHVVNLQCTACGHRGGYSSILVGGMISLISQWPMQASLQFOGY 239
Db 181 kvthalhsvyvergagsghvvnlqctacgrrggysrrvggmls1sqwpmqas1qfqgy 240
QY 240 HLCGGSVITPIMITAACVVDLKLRSWTIQVGLSLLDNPAPSHLVEKVYHRSKPK 299
Db 241 hlcggsvitpimitaahcvvdlylplkswtiqvglysvldipapshlvekvyhskpk 300
QY 300 RLGNDIALMALKAGPLTNEEMIOPVCLPNSSEENFPDGKVCWTSGWATEDGAGDASPVLNH 359
Db 301 rlgndialmalkagpltnemlqpcvclpnseenfpdgkvcwtswgatedg-gdaspvnh 359
QY 360 AAVPLTSNPKCHNRDVYGGIISPSMCLAGYIITGGVNSCQGSGGGPVCQERLWIKVAT 419
Db 360 aavplsnkicnhrdvyyggiispsmclagytggvdascqgdggpvcqerrlwikvat 419
QY 420 SFGIGCAEWNKPGVYTRVTSFLDWHQMERDIKT 454
Db 420 sfvigcaevnkpgvytrvtsfldwhqmerdikt 454

```

RESULT 5

AY72093 AAY72093 standard; Protein; 327 AA.

XX

AC AAY72093;

XX

DT 28-MAR-2001 (first entry)

XX

DE , Human serine protease #4 encoded by clone HCHAK72.

XX

KW Human; serine protease; osteopathic; immunosuppressive; antiallergic; antiinflammatory; cytostatic; cardiant; neuroprotective; nootropic; neuroleptic; vulnery; ophthalmological; antibacterial; antiviral; antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma; connective tissue disorder; autoimmune disorder; wound healing; asthma; systemic lupus erythematosus; male reproductive system disorder; testicular cancer; digestion and food absorption disorder; arrhythmia; Crohn's disease; neurodegenerative disease; Alzheimer's disease; allelic behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia; cardiovascular disorder; ocular disorder; drug screening.

XX

OS Homo sapiens.

XX

PN W020008247-A2.

XX

PD 16-NOV-2000.

XX

PF 05-MAY-2000; 2000WO-US12207.

XX

PR 07-MAY-1999; 99US-0133239.

PR 20-MAY-1999; 99US-0135163.

PR 03-AUG-1999; 99US-0147005.

PR 09-SEP-1999; 99US-0152935.

PR 01-NOV-1999; 99US-0162979.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Shi Y, Young PE, Ni J;

DR WPI; 2000-679799/66.

XX

DR N-PSDB; AAD02323.

XX

PT New nucleic acid molecules encoding human serine protease polypeptides useful for diagnosis, prevention and/or treatment of disorders e.g. osteoporosis, lupus erythematosus and Alzheimer's -

XX

PS Claim 12; Page 280-281; 289pp; English.

CC The present sequence is human serine protease #4 from clone HCHAK72 (ATCC Deposit No: PTA27).

the invention relates to nucleic acid sequences and cell lines claimed. It is used in methods for the diagnosis, prevention and treatment of various disorders related to serine protease such as bone formation disorders (osteoporosis), connective tissue disorders (arthritis), autoimmune disorders (systemic lupus erythematosus), wound healing, male reproductive system disorders (testicular cancer), digestion and food absorption disorders (Crohn's disease), neurodegenerative diseases (Alzheimer's disease), behavioral disorders (Roulette's syndrome), proliferative and cancerous conditions (acute myelogenous leukaemia), allergic reactions (asthma), cardiovascular disorders (arrhythmia), ocular disorders (glaucoma) and infectious diseases caused by bacteria, viruses, fungi or parasites. It is also useful for screening therapeutic compounds. Serine proteases are used as immunological probes or polymorphic markers for the identification of chromosomes, cells and tissues in biological samples. Identification of male contraceptive agents, delivery of compositions to targeted cells expressing a receptor for serine protease, hybridisation probes and molecular weight markers. Serine protease nucleic acids are also useful in gene therapy. Note: The present sequence shown in page 280-281 of sequence listing has been assigned SEQ ID NO: 14. But the sequence, human serine protease epidemic growth factor (EGF)-like domain (AYV2114) shown in page 12 is also referred as SEQ ID NO: 14.

PD	08-SEP-2000.	DT	08-FEB-2001 (first entry)
XX		XX	
PF	02-MAR-2000; 2000WO-US05612.	DE	Human cancer associated protein sequence SEQ ID NO:1017.
XX		XX	
PR	03-MAR-1999; 99US-0261416.	KW	Human; cancer associated gene; cancer; antigen; detection; cancer;
XX		KW	diagnosis; cytosolic; proliferative; vunlear; immunomodulator;
PA	(UZAR-) UNIV ARKANSAS.	KW	antidiabetic; antiallergic; antihistaminic; antirheumatic; antiarthritic; antiviral;
XX		KW	antiflammatory; antithyroid; antiinflammatory; antibacterial; cardiant;
PT	O'Brien TJ, Underwood LJ;	KW	dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
XX		KW	vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
DR	WPI; 2000-53263/48.	KW	immune disorder; haematopoietic cell disorder; autoimmune disorder;
DR	N-PSDB; AAC93843.	KW	allergic reaction; graft versus host disease; organ rejection;
XX		KW	haemostatic; thrombolytic; cardiovascular disorder; infection;
PT	DNA fragment encoding tumor associated differentially-expressed gene 12	KW	neurological disease; drug screening.
PT	protein used for diagnosing and treating malignant hyperplasia and	XX	
PT	cancers including ovarian cancer	XX	
PS	Claim 3: Page 80-81; 118pp; English.	OS	Homo sapiens.
XX		XX	
CC	This invention relates to a novel transmembrane serine protease called	PN	W020053350-A1.
CC	tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is	XX	
CC	located on chromosome 17. Sequences AAC93842-A93845 and AAB3246-B32249	PD	21-SEP-2000.
CC	represent human TADG-12 cDNA and their corresponding protein sequences.	XX	
CC	A splice variant of TADG-12 (TADG-12V) leads to a truncated protein	PF	08-MAR-2000; 2000WO-US05882.
CC	product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12	XX	
CC	exhibits cytostatic activity, and can be used in vaccines and in gene	PR	12-MAR-1999; 99US-0124270.
CC	therapy. TADG-12 nucleotide and protein sequences are used in the	XX	
CC	diagnosis of malignant hyperplasia and cancers of the ovary, breast,	PA	(HUMA-) HUMAN GENOME SCI INC.
CC	lung, colon, prostate and other cancers where TADG-12 is overexpressed.	XX	
CC	TADG-12 is particularly used as tumour marker for early disease	PT	Rosen CA, Rubin SM;
CC	diagnosis. TADG12 proteins or fragments can be used to vaccinate an	XX	
CC	individual with cancer suspected of having a cancer or at risk of	DR	DR
CC	getting cancer. Sequences AAC93846-A93853 represent PCR primers used for	WPI	WPI; 2000-587533/55.
CC	amplifying the TADG-12 cDNA sequence, and in the quantitative analysis of	XX	N-PSDB; AAC77781.
CC	TADG-12 mRNA. AAB3250 represents a peptide fragment of TADG-12, used to	CC	
CC	create anti-TADG-12 antibodies. Sequences RAB3251-B32369 represent	CC	
CC	TADG-12 peptides which target HLA, and may be used in a vaccine or for	CC	
CC	immune stimulation.	CC	
XX	Sequence 294 AA;	CC	
SQ	Query Match 55.3%; Score 1354; DB 21; Length 294;	CC	AAC77607 to AAC78448 encode the human cancer associated proteins given
CC	Best Local Similarity 88.1%; Pred. No. 7-8e-10; Mismatches 20; Indels 8; Gaps 2;	CC	in AAB43398 to AAB4239. The proteins can have activities based on the
CC	Matches 258; Conservative 7; Mismatches 20; Indels 8; Gaps 2;	CC	tissues and cells the genes are expressed in. Example of activities
CC	include: cytostatic; proliferative; vulnerability; immunomodulator;	CC	antidiabetic; antiasthmatic; antiarthritic;
CC	antinflammatory; antithyroid; antiallergic; antibacterial; antiviral;	CC	dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC	antitoxic; antipsoriatic and antianalgesic. The	CC	nootropic; vasotropic; antipsoriatic and antianalgesic. The
CC	polynucleotides and polypeptides can be used for preventing, treating or	CC	ameliorating medical conditions and diagnosing pathological conditions.
CC	Polynucleotides, polypeptides, antibodies, agonists and antagonists from	CC	the present invention may be used to treat immune disorders by activating
CC	or inhibiting the proliferation, differentiation or mobilisation of	CC	immune cells, to treat disorders of haematopoietic cells, autoimmune
CC	disorders, allergic reactions, graft versus host disease and organ	CC	disorders, modulating haemostatic or thrombolytic activity, modulate
CC	inflammation, cancers, cardiovascular disorders, neurological disease and	CC	bacterial or viral infections. The peptides, nucleotides, antibodies,
CC	agonists and antagonists may be also be used in drug screens. AAC78449 to	CC	AAC78457 and AAB4240 represent sequences used in the exemplification of
CC	the present invention.	CC	the present invention.
SQ	Sequence 248 AA;	XX	
Query Match 54.7%; Score 1339; DB 21; Length 248;	XX		
Best Local Similarity 100.0%; Pred. No. 1e-99; Mismatches 0; Indels 0; Gaps 0;	XX		
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX		
Db	241 hlgggsvitpwlwitaahcveyevapreradrrgrkllcwtkptmkgrprph 293	QY	210 RGGSSRTGGGNMNLSSQWPWQASLQLQGYHLLGGSVITPWLWITAACVYDYLPSWT 269
RESULT 7		Db	4 rryyssrrggggmnlssqwpwqasllqfcgqyhgccgsvitpwlwitaahcveydylpswt 63
AAB43572		QY	270 IQGLVLSIDNPAPSHLVEKIVHHSKPKRGGNDIALMKGAGPLTNEEMTOPVCLNSE 329
ID AAB43572 standard; Protein: 248 AA.		Db	64 iqvgivslldinpapshlvekivhskpkrgndialmklaqpitinemqpcipns 123
XX			

Query Match 42.3%; Score 1034; DB 22; Length 193;
 Best Local Similarity 100.0%; Pred. No. 2.2e-75;
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 ENPPDGKVWTSGGWATEDGAGDASPVLNHAAPLISNKICNRDVYGGIISPMCLAGY 389
																			; Sequence 193 AA;																																																	

RESULT 8
 ID AAG73966 standard; Protein: 193 AA.
 AC AAG73966;
 XX DR 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:4730.
 KW colorectal carcinoma.
 XX Homo sapiens.

PR WO200122920-A2.

PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

PT Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI; 2001-235357-24.
 DR N-PSDB; AAAH3337.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers.

XX PS Claim 11; Page 6530-6531; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in active proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated P's, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB7789 represent sequences used in the exemplification of the present invention.

CC N, P, Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for N-PSDB ID No:1027 to 1052, 7921 and 7922.

XX Sequence 193 AA;

RESULT 9
 ID AAY92050 standard; Protein: 492 AA.
 AC AAY92050;
 XX DT 01-AUG-2000 (first entry)

DE HRPcA6/7 polypeptide from androgen-inducible gene clone.

XX KW Androgen inducible; testosterone; prostate cancer; cytostatic; TMPRSS2; diagnosis.

XX OS Homo sapiens.

XX PN WO200018961-A2.

PR 30-SEP-1998; 98US-0163759.

XX PR 30-SEP-1998; 98US-0164159.

XX PA (MILL-) MILLENTIUM PHARM INC.

PT Macbeth KJ, Shyjan AW;

XX DR WPI; 2000-29182/25.

XX DR N-PSDB; AAA08803.

PT Novel methods for identifying compounds for treating prostate cancer comprising measuring the level of expression or activity of 1 or more of 11 genes or their products.

XX PS Claim 2; Fig 3; 108pp; English.

CC This protein is encoded by a gene which is androgen (e.g. testosterone) inducible in androgen-dependent prostate cancer cells (e.g. LNCaP cells) and constitutively expressed in androgen-independent prostate cancer cells (e.g. LN3 LNCaP cells). Agents which decrease the expression or activity of these clones may slow or arrest the growth of prostate cancer cells or may kill them. HRPcA6/7 can be obtained from the sequence of the known gene for TMPRSS2. A compound useful for treating prostate cancer can be identified in a novel method comprising measuring the expression level, or activity, of HRPcA2, 3, 6/7, 8, 9, 10, 13, 14, 15, 19, or peripheral-type benzodiazepine receptor (PR) in a cell, the presence and absence of a test compound. The sequences may also be used in diagnosis of prostate cancer and to determine efficacy of treatment for prostate cancer.

Query Match	35	88;	Score	876.5;	DB	21;	Length	492;
Best Local Similarity	45.9%	Pred.	No.	2.9e62;				
Matches	189;	Conservative	53;	Mismatches	149;	Delets	21;	Gaps
XX								
QY	52	IVIGVALLALAGLGIN--DCSGK-YRCRSCFKCIELIARDGVSICKDGDDBYRC	107					
Db	89	ltlgftflvgalaaglwlwgmskcsnsqiecdssgtcinspnwcdgvshcpgeedenc	148					
QY	108	VRYGGONAVLQVFAA-ASWKTMCSDDWGHYANVACAGLGFPSYVSSINLRYSSLE3Q	164					
Db	149	vrlqygnfnilqvyssqrkswhpvcqddwnenygracromgkymnyqgivdsgsts	208					
QY	165	FREEFVSDHLLPDKVYALHHSYVRECGASGHVWTCTACGHRGGS--SRVGGNM	222					
Db	209	fmkintsaqng--diyklhhs---dacsavksvslciacygnlnsrsqslvsges	261					
QY	223	SLSQWMPQASLQPGQYHCGGSITPLHILARCV-YDLILKSWTQGLV--SLD	279					
Db	262	alpgawpwpqyvslvqnvhccggsitpeviwtahcveplnppwhatafirlqsfm	321					
QY	280	NPAPSHLVLVKVIVHSKYKRRRLGNDIALMKLAGPLFENEMIQPVYCLPNSEBNFPDGKVCW	339					
Db	322	yga qyqekvishpnydsktkndialmklpflnolvpolpnmqlqpeqlw	380					
QY	340	TSGWATEGAGGASPAPVNHAAVPLISNICKNHRDVYGGTISPSMCLAGSLTGGSOCG	399					
Db	381	iswgwatec-kgtktsevlnaakvlietrcntryvduhltpanicaqflqgnvdsccq	439					
QY	400	DSGGPLWCGCERRRLWIKLWGTSGFCGAENKPGYTRVSLFDLWHEQMRD	451					
Db	440	dsggplvtksnniwlglidtswgsgcakayrpgyygnvftdwlyrqmrad	491					
RESULT	10							
ID	AAM01315							
XX	-	AAM01315 standard; Protein; 492 AA.						
AC	AAM01315;							
XX								
DT	04-OCT-2001	(first entry)						
XX								
DE	P1000C amino acid sequence.							
XX								
KW	Human; prostate cancer; prostate-specific; diagnosis; vaccine;							
XX								
KW	cytostatic; gene therapy; metastasis.							
XX								
OS	Home sapiens.							
XX								
PN	WO200151633-A2.							
XX								
PD	19-JUL-2001.							
XX								
PF	16-JAN-2001; 2001WO-US01574.							
XX								
PR	14-JAN-2000; 2000US-0483672.							
XX								
PA	(CORI-) CORIXA CORP.							
XX								
PI	Xu J, Dillon DC, Mitcham JL, Harlocker ST, Jiang Y, Reed SG; Kalcs MD, Farmer GR, Day CH, Retter MW, Stolk JA, Skeky YAW; Wang A, Meagher MJ;							
XX								
DR	WPI; 2001-425873/45.							
XX								
PT	New poly nucleotide encoding a prostate-specific protein, for diagnosing, monitoring and treating prostate cancer in a patient and for use in vaccines -							
XX								
PS	Claim 2; Page 539-540; 543pp; English.							
XX								
CC	The present invention describes poly nucleotide sequences (I) which encode							

CC prostate-specific proteins (III), (I) and (III) have cytostatic activity, and can be used in vaccine production and gene therapy. (I), (II), and antibodies to (II), fusion proteins comprising (II), and isolated T cells prepared using (I) or (II) are used to treat cancer in a patient. (I) and the antibodies are also used in the detection of cancer in a patient. The cancer that is diagnosed or treated is particularly prostate cancer. (I) and (III) can be used in vaccines. The antibodies or (I) can be used for monitoring the progression of cancer in a patient. (I) and (III) can also be used to improve diagnostic and therapeutic methods for prostate cancer. They can indicate the level of metastasis as well as the prostate volume. AAM93357 to AAM93944 and AAM01115 to AAM01318 represent polynucleotide and amino acid sequences used in the exemplification of the present invention.

PA (MYRI-) MYRIAD GENETICS INC.
 XX
 PI Wong AKC, Tavtigian SV, Teng DHF;
 XX
 DR WPI; 2000-170914/15.
 XX
 DR N-PSDB; AAZ87786.
 XX
 PT Novel tumor suppressor TMPRSS2 used for the diagnosis and prognosis of
 human cancer -
 XX
 PS Claim 55; Page 77-79; 89pp; English.
 XX
 CC The invention provides a new tumour suppressor gene, designated TMPRSS2.
 CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for
 CC diagnosing and pronouncing predisposition to cancer in humans. The
 CC polypeptides may also be used in assays to screen for compounds with
 CC anti-cancer or therapeutic properties.
 CC The polypeptides are also useful
 CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides
 CC may be used for gene therapy and protein therapy. The present sequence
 XX represents the TMPRSS2 polypeptide.

Sequence 492 AA;

Query Match 35.8%; Score 875.5; DB 21; Length 492;

Best Local Similarity 45.9%; Pred. No. 3.5e-62; Matches 189; Conservative 53; Mismatches 149; Indels 21; Gaps 11;

Qy 52 I V I G I T A L I A L A G I G I H R - - D C S G K - Y C R C R S F K C T E L I A R C D G V S D C K D G E D E Y R C 107
 Db 89 I t l g f t f l v g a a l a g l l w k m g s k o s n g i e c d s g t c i n p s n w d g v s h c p g g e d e n r c 148
 Qy 108 V R V G Q N A V I Q V F T A - - A S K T M C S D D W K C H Y A N V A C A O Q G F P - S V V S D N L R V S L E Q 164
 Db 149 v r l y g p n f l l q v y s s q r k s w h p v c q d d w n e n y g r a c r d g y k n f y s s q g i v d s g s t s 208
 Qy 165 F R E E V S I D H L P D D K V T A I H S V V R E C A S G H V T L O C T A C C H R R G S - - S R I V G G M 222
 Db 209 f m k l n t s a g n - - d i y k k i y s - - d a c s k a v s l r i c a c g v n l n s r s q s r i v g g e s 261

Qy 223 S L L S O W P W Q A S L Q R Q G Y H C G G S V I T P L W I T A H C V - D L Y L P K S W T I Q V G L V - S L L D 279
 Db 262 a l p g a w p q v s l h v q n v h v e c g s s i t p e w i v t a h c v e k p l n p w h w t a f a g i r q s f m f 321
 Qy 280 N P A P S H L V E K I V Y H S K Y K P R R G N D I A L M K L A G P L T F N E M I Q P V C L P N S E E N F D G K V C W 339
 Db 322 y g a - g y q v e k v i s h p n y d s k t k n d i a l m k l q p l t f n d l v k p c l p n p g m l q p e q l c w 380
 Qy 340 T S G W G A T E D G A D S P V L N H A V P L I S N K T C I N H R D V Y G G I S P E M C A G Y L T G V D S C G 399
 Db 381 l s g w g a t e e - k g k t s e v l n a k v l i l e t g r o n s v y d h l p a n i c a g f l q g n v d s c g 439
 Qy 400 D S G G P L V C O B R R L K L V G A T S F C G C A E N K P G V Y T R V S F L D M I H E Q M R D 451
 Db 440 d s g g p l v t s k n n i w l i g a t s w g g c a k a y r p g y y g n v n v f t d v i r q p r a d 491

Query Match 35.8%; Score 875.5; DB 21; Length 492;
 Best Local Similarity 45.9%; Pred. No. 3.5e-62; Matches 189; Conservative 53; Mismatches 149; Indels 21; Gaps 11;

Qy 52 I V I G I T A L I A L A G I G I H R - - D C S G K - Y C R C R S F K C T E L I A R C D G V S D C K D G E D E Y R C 107
 Db 89 I t l g f t f l v g a a l a g l l w k m g s k o s n g i e c d s g t c i n p s n w d g v s h c p g g e d e n r c 148
 Qy 108 V R V G Q N A V I Q V F T A - - A S K T M C S D D W K C H Y A N V A C A O Q G F P - S V V S D N L R V S L E Q 164
 Db 149 v r l y g p n f l l q v y s s q r k s w h p v c q d d w n e n y g r a c r d g y k n f y s s q g i v d s g s t s 208
 Qy 165 F R E E V S I D H L P D D K V T A I H S V V R E C A S G H V T L O C T A C C H R R G S - - S R I V G G M 222
 Db 209 f m k l n t s a g n - - d i y k k i y s - - d a c s k a v s l r i c a c g v n l n s r s q s r i v g g e s 261

Qy 223 S L L S O W P W Q A S L Q R Q G Y H C G G S V I T P L W I T A H C V - D L Y L P K S W T I Q V G L V - S L L D 279
 Db 262 a l p g a w p q v s l h v q n v h v e c g s s i t p e w i v t a h c v e k p l n p w h w t a f a g i r q s f m f 321
 Qy 280 N P A P S H L V E K I V Y H S K Y K P R R G N D I A L M K L A G P L T F N E M I Q P V C L P N S E E N F D G K V C W 339
 Db 322 y g a - g y q v e k v i s h p n y d s k t k n d i a l m k l q p l t f n d l v k p c l p n p g m l q p e q l c w 380
 Qy 340 T S G W G A T E D G A D S P V L N H A V P L I S N K T C I N H R D V Y G G I S P E M C A G Y L T G V D S C G 399
 Db 381 l s g w g a t e e - k g k t s e v l n a k v l i l e t g r o n s v y d h l p a n i c a g f l q g n v d s c g 439
 Qy 400 D S G G P L V C O B R R L K L V G A T S F C G C A E N K P G V Y T R V S F L D M I H E Q M R D 451
 Db 440 d s g g p l v t s k n n i w l i g a t s w g g c a k a y r p g y y g n v n v f t d v i r q p r a d 491

Query Match 35.8%; Score 875.5; DB 21; Length 492;
 Best Local Similarity 45.9%; Pred. No. 3.5e-62; Matches 189; Conservative 53; Mismatches 149; Indels 21; Gaps 11;

Qy 52 I V I G I T A L I A L A G I G I H R - - D C S G K - Y C R C R S F K C T E L I A R C D G V S D C K D G E D E Y R C 107
 Db 89 I t l g f t f l v g a a l a g l l w k m g s k o s n g i e c d s g t c i n p s n w d g v s h c p g g e d e n r c 148
 Qy 108 V R V G Q N A V I Q V F T A - - A S K T M C S D D W K C H Y A N V A C A O Q G F P - S V V S D N L R V S L E Q 164
 Db 149 v r l y g p n f l l q v y s s q r k s w h p v c q d d w n e n y g r a c r d g y k n f y s s q g i v d s g s t s 208
 Qy 165 F R E E V S I D H L P D D K V T A I H S V V R E C A S G H V T L O C T A C C H R R G S - - S R I V G G M 222
 Db 209 f m k l n t s a g n - - d i y k k i y s - - d a c s k a v s l r i c a c g v n l n s r s q s r i v g g e s 261

Qy 223 S L L S O W P W Q A S L Q R Q G Y H C G G S V I T P L W I T A H C V - D L Y L P K S W T I Q V G L V - S L L D 279
 Db 262 a l p g a w p q v s l h v q n v h v e c g s s i t p e w i v t a h c v e k p l n p w h w t a f a g i r q s f m f 321
 Qy 280 N P A P S H L V E K I V Y H S K Y K P R R G N D I A L M K L A G P L T F N E M I Q P V C L P N S E E N F D G K V C W 339
 Db 322 y g a - g y q v e k v i s h p n y d s k t k n d i a l m k l q p l t f n d l v k p c l p n p g m l q p e q l c w 380
 Qy 340 T S G W G A T E D G A D S P V L N H A V P L I S N K T C I N H R D V Y G G I S P E M C A G Y L T G V D S C G 399
 Db 381 l s g w g a t e e - k g k t s e v l n a k v l i l e t g r o n s v y d h l p a n i c a g f l q g n v d s c g 439
 Qy 400 D S G G P L V C O B R R L K L V G A T S F C G C A E N K P G V Y T R V S F L D M I H E Q M R D 451
 Db 440 d s g g p l v t s k n n i w l i g a t s w g g c a k a y r p g y y g n v n v f t d v i r q p r a d 491

RESULT 13	AAM01317	ID AAM01317 standard; Protein; 393 AA..	Db 174 lhvqvhvbcgssittpewitaacvvekplnpnphntafqilrqsfmfyga-qyqvkv
XX	XX	XX	Qy 291 VYHSKYKPRKLGNIDLALMKGAPLTFNEMTQPVCLPNSENFPDGRKVCWMSGMGRATEGDA 350
AC	AC	AC	Db 233 ishpdydkstknnaialmlqkptfndkvkpcplnpqam1lpeqlcwisqgatee-k 291
XX	XX	XX	DT 04-OCT-2001 (first entry)
DE	DE	DE	P1000C partial amino acid sequence 100-492.
KW	KW	KW	Human; prostate cancer; prostate-specific; diagnosis; vaccine; cytostatic; gene therapy; metastasis.
OS	OS	OS	Homo sapiens.
XX	XX	XX	PN WO200151633-A2.
XX	XX	XX	PD 19-JUL-2001.
PF	PF	PF	PR 16-JAN-2001; 2001WO-US01574.
XX	XX	XX	PR 14-JAN-2000; 2000US-0483672.
PA	PA	PA	(CORI-) CORIXA CORP.
XX	XX	XX	PR Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG; PI Kalos MD, Ranger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW; PI Wang A, Meagher MJ;
XX	XX	XX	DR WPI; 2001-425873/45.
XX	XX	XX	PT New polynucleotide encoding a prostate-specific protein, for diagnosing, monitoring and treating prostate cancer in a patient and for use in vaccines -
XX	XX	XX	PS Claim 2; Page 540-541; 543pp; English.
CC	CC	CC	The present invention describes polynucleotide sequences (I) which encode prostate-specific proteins (II). (I) and (II) have cytostatic activity, and can be used in vaccine production and gene therapy. (I), (II), antibodies to (II), fusion proteins comprising (II), and isolated T cells prepared using (I) or (II) are used treat cancer in a patient. (I) and the antibodies are also used in the detection of cancer in a patient. The cancer that is diagnosed or treated is particularly prostate cancer. (I) and (II) can be used in vaccines. The antibodies or (I) can be used for monitoring the progression of cancer in a patient. (I) and (II) can also be used to improve diagnostic and therapeutic methods for prostate cancer. They can indicate the level of metastasis as well as the prostate volume. AAM01317 to AAM013944 and AAM01115 to AAM01318 represent polynucleotide and amino acid sequences used in the exemplification of the present invention.
CC	CC	CC	CC Sequence 393 AA;
XX	XX	XX	PS Claim 63; Page 86-88; 121pp; English.
CC	CC	CC	The present invention relates to prostate specific androgen regulated proteins. The invention may be used to determine an expression level of the prostate-specific proteins ARSP1; TMPRSS2, or PARM-1 in a fluid sample or prostate cell sample from an individual. It may also be used for diagnosing and predicting the susceptibility of a prostate neoplastic condition in an individual. Inhibitors of the proteins are useful for treating or preventing the progression of a prostate neoplastic condition.
CC	CC	CC	CC Sequence 492 AA;
QY 176	QY 176	QY 176	Query Match 35.5%; Score 868.5; DB 22; Length 393; Best Local Similarity 46.6%; Pred. No. 9.7e-62; Matches 187; Conservative 50; Mismatches 143; Indels 21; Gaps 11;
Db 121	Db 121	Db 121	Query Match 35.3%; Score 864.5; DB 21; Length 492; Best Local Similarity 45.4%; Pred. No. 2.6e-61; Matches 186; Conservative 55; Mismatches 148; Indels 21; Gaps 11;
QY 234	QY 234	QY 234	Query Match 35.3%; Score 864.5; DB 21; Length 492; Best Local Similarity 45.4%; Pred. No. 2.6e-61; Matches 186; Conservative 55; Mismatches 148; Indels 21; Gaps 11;

with a cytotoxic agent are useful for treating cancer. The present sequence represents a ovrial homolog protein, that can be used for the detection of the various cancers.

Thu Apr 11 09:27:12 2002

us-09-846-512-12.rag

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

April 11, 2002, 08:52:32 ; Search time 28.67 Seconds

(without alignments)

2316.276 Million cell updates/sec

Title:

US-09-846-512-12

Perfect score:

2447 1 MGENDPPAVEAPFSRSLFG. TRVTSLFDWIHQMERDLKT 454

Sequence:

Q9XZM7

Scoring table:

BLOSUM62

GapP 10.0 , GapExt 0.5

Searched:

473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters:

473505

Minimum DB seq length:

0

Maximum DB seq length:

2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPPREMBL_17:*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mic:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rat:*

12: sp_virus:*

13: sp_vertebrat:*

14: sp_unclassified:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT

1

Q9BXX1 ID PRELIMINARY;

AC Q9BXX1;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE ANDROGEN-REGULATED SERINE PROTEASE TMPRSS2..

GN

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.

RN NCBI_TAXID=9605;

RN

RP SEQUENCE FROM N.A.

RX MEDLINE:2113912; Pubmed=11245484;

RA Afar D.E.H., Vivanco I., Hubert R.S., Kuo J., Chen E., Saffran D.C.,

RA Raitano A.B., Jakobovits A.;

RT "Catalytic cleavage of the androgen-regulated TMPRSS2 protease results

RT in its secretion by prostate and prostate cancer epithelia.";

RL Cancer Res. 61:1686-1692(2001).

RN

RP SEQUENCE FROM N.A.

RA Mitchell S.C., Hubert R.S., Afar D.E.H.,

RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL:AF270487; AAK29280.1; -

KW

PROTEASE.

FT CHAIN 255 492 ANDROGEN-REGULATED SERINE PROTEASE

FT

SO SEQUENCE 492 AA; 53859 MW; C05B5531C8A311C7 CRC64;

Query Match Similarity 35.8%; Score 875.5; DB 4; Length 492;

Matches 189; Conservative 53; Mismatches 149; Indels 21; Gaps 11;

O9XZM7

20	527	21.5	279	11	Q9QZ74	Q9QZ74 rattus norv
21	512.5	20.9	310	11	Q9QZ74	Q9QZ74 mus musculu
22	512	20.9	422	4	Q9UL52	Q9UL52 homo sapien
23	511	20.9	581	5	Q9XZM7	Q9XZM7 stronglyloce
24	510.5	20.9	812	11	Q9R03	Q9R03 ratus norv
25	508.5	20.8	389	13	Q9PVX7	Q9PVX7 xenopus lae
26	507	20.7	311	11	Q9QUL7	Q9QUL7 mus musculu
27	506.5	20.7	624	11	Q9DATA3	Q9DATA3 ovis aries
28	505.5	20.7	810	4	Q15146	Q15146 homo sapien
29	500	20.4	806	6	Q18783	Q18783 macroscopus eu
30	499.5	20.4	329	6	Q9GL10	Q9GL10 ovis aries
31	499.5	20.4	334	6	Q46507	Q46507 papio hamad
32	499.5	20.4	339	11	Q99JCB	Q99JCB ovis aries
33	496.5	20.3	273	6	Q9XSM1	Q9XSM1 ovis aries
34	496.5	20.3	317	13	Q9DGR3	Q9DGR3 xenopus lae
35	494.5	20.2	325	5	Q15944	Q15944 sarcophaga
36	494.5	20.2	571	4	Q9Y495	Q9Y495 homo sapien
37	494.5	20.2	749	11	Q99JCB	Q99JCB ovis aries
38	487.5	19.9	273	6	Q9XSM1	Q9XSM1 ovis aries
39	484.5	19.8	454	6	Q46506	Q46506 papio hamad
40	484	19.8	502	11	Q9CW97	Q9CW97 mus musculu
41	483.5	19.8	282	11	Q9D4T3	Q9D4T3 mus musculu
42	481.5	19.7	421	11	Q60491	Q60491 capra porce
43	480.5	19.6	275	4	Q15661	Q15661 homo sapien
44	479.5	19.6	260	13	Q9W7Q3	Q9W7Q3 paralichthys
45	479	19.6	264	11	Q9EQ28	Q9EQ28 rattus norv

RESULT 4

Q9BYEL PRELIMINARY; PRT; 537 AA.

AC 09BYEL; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE MOSATIC SERINE PROTEASE.

GN MSPC.

OS Homo sapiens (Human)

OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Cetartiodactyla; Suina; Suidae; Sus.

OC Mammalia; Butheria; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1] SEQUENCE FROM N A.

RC TISSUE=PLASMA;

RA Takahashi T., Kimura A., Okimura H., Hanabata T.;

RP "Porcine liver Plasma kallikrein.";

RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY (S1).

DT CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).

DR EMBL: AB02425; BAA37147.1; -.

DR HSSP: P00766; ICHG.

DR MEROPS: S01.212; -.

DR MEROPE: IPR00177; APPLE.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001314; PAN.

DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00024; PAN; 4.

DR Prints: PR00005; APPLEDOMAIN.

DR Prints: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; TIP_SPC; 1.

DR PROSITE: PS00495; APPLE; 4.

DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.

KW Hydrolase; Serine protease.

SEQUENCE 537 AA; 58102 MW; A395F4E8816DABCF CRC64;

Query Match 29.1%; Score 712.5; DB 4; Length 537;

Best Local Similarity 38.9%; Pred. No. 1.5e-58;

Matches 145; Conservative 64; Mismatches 147; Indels 17; Gaps 7;

QY 85 CIELTARCDGVSCKDGEDEYRCVRYGGONAVLQVFTAAS--WKKMCSDWKGHVNAYVAC 142

Db 174 CPKHAVRCDGWDCKLKSDELGCWRDWDKSLKITYSGSSHQWLPICTCSSSHWNDSSEKTC 233

QY 143 AQLGPSPVSYSDNLNRYVSSLEGQFREFVSDHLLDDKVKYALHHVSYVREGCASHVWLT 202

Db 234 RQLGFSRASHRHTEVA-----HRDFANSFSILRYN--STIODESLH-RSHCPSORYISL 282

QY 203 OCTACCHRRGYSRIVGGNNMSSLQSPWQASLQFOGYHLCGGSVITPLWITAAHCYVYDYL 262

Db 283 QCSHGG-LRAMTGRIVGALASDKRKPWQVSLHFGTHICGTLIDAQWVLTAAHCFFVT 341

QY 263 --YLPKSWTIVQGYLSSLDNPAPSHLVEKIVYHSKYKPKLGNIDALMKLAGPCTFNEMI 320

Db 342 REKVLGKVKYTAGTSNLHOLEPAAIE-IIINSNTDEDDYDIALMRLSKPLTSI 400

QY 321 QPVCLPNSNEFPDKVCMWSGWAGTEDGAGDASVNLHAAVPLJNSNKTCNRHRYGGII 380

Db 401 RPACLPMHGPQFSLNBTWCWTFGKTRDDKTSPLFREVQVNLLDFKKCNDYLVYDSYL 460

QY 381 SPSMLCAGYLTTGGWISQCQGDGGPLVYQERRLWLYGATSGFGIGCAEVNPKGVYTRVTSF 440

Db 461 TPRMAGCAGDHLGGBRQSCQGDGGPLVYQERRLWLYGATSGFGIGCAEVNPKGVYTRVTSF 520

QY 441 LDWIIHQMEROLK 453

Db 521 LPWIVSKMSEVR 533

RESULT 5

Q9DB10 PRELIMINARY; PRT; 799 AA.

AC 09DB10; 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

RC TISSUE=PLASMA;

RA Takahashi T., Kimura A., Okimura H., Hanabata T.;

RP "Porcine liver Plasma kallikrein.";

RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY (S1).

DT CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).

DR EMBL: AB02425; BAA37147.1; -.

DR HSSP: P00766; ICHG.

DR MEROPS: S01.212; -.

DR InterPro: IPR00177; APPLE.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001314; PAN.

DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00024; PAN; 4.

DR Prints: PR00005; APPLEDOMAIN.

DR Prints: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; TIP_SPC; 1.

DR PROSITE: PS00495; APPLE; 4.

DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.

KW Hydrolase; Serine protease.

SEQUENCE 643 AA; 72227 MW; AFEF923E3CB0A CRC64;

Query Match 23.7%; Score 579; DB 6; Length 643;

Best Local Similarity 36.9%; Pred. No. 7.4e-16;

Matches 128; Conservative 55; Mismatches 118; Indels 46; Gaps 10;

QY 132 DWKGHVNAYVACQALGPSPVSYSDNLNRYVSSLEGQFREFVSD-----HII 176

Db 309 DFEGEELNV-----TFVOGANICQETCTKTRCIRQFTYSLHPPDCRCBCKCSLRS 360

QY 177 PDDKVYALHHVSYVREG----CASCHVWVQCTAGHRRGQSSRIVGGNMSLSSQWPO 231

Db 361 SDGSPKTRITHGMRASSGYSLRLCRSG----DHSACATKA--NTRIVGGDSFGEWPWQ 413

QY 232 ASLQFO---GHLCGGSVITPLWITAAHCYVYDYLIPKSWTIVQVGI--VSLLDNFAPSHL 286

Db 414 VSLQAKLRAQNHLCGGSIIQHQLVLTAAHCFFGSLSPDIKRIYIGGLINNSEITKETPFSQ 473

QY 287 VEVTKVHSKYKPKLGNIDALMKLAGPCTFNEMIOPVCLNSENFPDGKVCWISGAGT 346

Db 474 VKEIITHQVKYKILESGHDLKLEPLNTYDFQPKICLSSRDTNVVYTCWVWGF 533

QY 347 EDGAGDASVNLHAAVPLJNSNKTCNRHRYGGII 404

Db 534 EE-KGETIOTILQKVNIPLVNEECOKSYRD--HKISKOMICAGYREGGDKACKGESEGGP 589

QY 405 LVQERRLWLYGATSGFGIGCAEVNPKGVYTRVTSF LDWIIHQMEROLK 451

Db 590 LUCKYNGIWHVGGTISWEGSCARREOPGVYTKVIEWMDWLEKTD 636

RESULT 6

Q9DB10 PRELIMINARY; PRT; 799 AA.

AC 09DB10; 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

RC TISSUE=PLASMA;

RA Takahashi T., Kimura A., Okimura H., Hanabata T.;

RP "Porcine liver Plasma kallikrein.";

RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY (S1).

DT CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).

DR EMBL: AB02425; BAA37147.1; -.

DR HSSP: P00766; ICHG.

DR MEROPS: S01.212; -.

DR InterPro: IPR00177; APPLE.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001314; PAN.

DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00024; PAN; 4.

DR Prints: PR00005; APPLEDOMAIN.

DR Prints: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; TIP_SPC; 1.

DR PROSITE: PS00495; APPLE; 4.

DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.

KW Hydrolase; Serine protease.

SEQUENCE 643 AA; 72227 MW; AFEF923E3CB0A CRC64;

Query Match 23.7%; Score 579; DB 6; Length 643;

Best Local Similarity 36.9%; Pred. No. 7.4e-16;

Matches 128; Conservative 55; Mismatches 118; Indels 46; Gaps 10;

QY 132 DWKGHVNAYVACQALGPSPVSYSDNLNRYVSSLEGQFREFVSD-----HII 176

Db 309 DFEGEELNV-----TFVOGANICQETCTKTRCIRQFTYSLHPPDCRCBCKCSLRS 360

QY 177 PDDKVYALHHVSYVREG----CASCHVWVQCTAGHRRGQSSRIVGGNMSLSSQWPO 231

Db 361 SDGSPKTRITHGMRASSGYSLRLCRSG----DHSACATKA--NTRIVGGDSFGEWPWQ 413

QY 232 ASLQFO---GHLCGGSVITPLWITAAHCYVYDYLIPKSWTIVQVGI--VSLLDNFAPSHL 286

Db 414 VSLQAKLRAQNHLCGGSIIQHQLVLTAAHCFFGSLSPDIKRIYIGGLINNSEITKETPFSQ 473

QY 287 VEVTKVHSKYKPKLGNIDALMKLAGPCTFNEMIOPVCLNSENFPDGKVCWISGAGT 346

Db 474 VKEIITHQVKYKILESGHDLKLEPLNTYDFQPKICLSSRDTNVVYTCWVWGF 533

QY 347 EDGAGDASVNLHAAVPLJNSNKTCNRHRYGGII 404

Db 534 EE-KGETIOTILQKVNIPLVNEECOKSYRD--HKISKOMICAGYREGGDKACKGESEGGP 589

QY 405 LVQERRLWLYGATSGFGIGCAEVNPKGVYTRVTSF LDWIIHQMEROLK 451

Db 590 LUCKYNGIWHVGGTISWEGSCARREOPGVYTKVIEWMDWLEKTD 636

RESULT 6

Q9DB10 PRELIMINARY; PRT; 799 AA.

AC 09DB10; 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

RC TISSUE=PLASMA;

RA Takahashi T., Kimura A., Okimura H., Hanabata T.;

RP "Porcine liver Plasma kallikrein.";

RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY (S1).

DT CC -!- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).

DR EMBL: AB02425; BAA37147.1; -.

DR HSSP: P00766; ICHG.

DR MEROPS: S01.212; -.

DR InterPro: IPR00177; APPLE.

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001314; PAN.

DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00024; PAN; 4.

DR Prints: PR00005; APPLEDOMAIN.

DR Prints: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; TIP_SPC; 1.

DR PROSITE: PS00495; APPLE; 4.

DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.

KW Hydrolase; Serine protease.

SEQUENCE 643 AA; 72227 MW; AFEF923E3CB0A CRC64;

Query Match 23.7%; Score 579; DB 6; Length 643;

Best Local Similarity 36.9%; Pred. No. 7.4e-16;

Matches 128; Conservative 55; Mismatches 118; Indels 46; Gaps 10;

QY 132 DWKGHVNAYVACQALGPSPVSYSDNLNRYVSSLEGQFREFVSD-----HII 176

Db 309 DFEGEELNV-----TFVOGANICQETCTKTRCIRQFTYSLHPPDCRCBCKCSLRS 360

QY 177 PDDKVYALHHVSYVREG----CASCHVWVQCTAGHRRGQSSRIVGGNMSLSSQWPO 231

Db 361 SDGSPKTRITHGMRASSGYSLRLCRSG----DHSACATKA--NTRIVGGDSFGEWPWQ 413

QY 232 ASLQFO---GHLCGGSVITPLWITAAHCYVYDYLIPKSWTIVQVGI--VSLLDNFAPSHL 286

Db 414 VSLQAKLRAQNHLCGGSIIQHQLVLTAAHCFFGSLSPDIKRIYIGGLINNSEITKETPFSQ 473

QY 287 VEVTKVHSKYKPKLGNIDALMKLAGPCTFNEMIOPVCLNSENFPDGKVCWISGAGT 346

Db 474 VKEIITHQVKYKILESGHDLKLEPLNTYDFQPKICLSSRDTNVVYTCWVWGF 533

QY 347 EDGAGDASVNLHAAVPLJNSNKTCNRHRYGGII 404

Db 534 EE-KGETIOTILQKVNIPLVNEECOKSYRD--HKISKOMICAGYREGGDKACKGESEGGP 589

QY 405 LVQERRLWLYGATSGFGIGCAEVNPKGVYTRVTSF LDWIIHQMEROLK 451

Db 590 LUCKYNGIWHVGGTISWEGSCARREOPGVYTKVIEWMDWLEKTD 636

Best local Similarity	31.3%	Pred.	No. 2.6e-44;	Score	564.5;	DB	4;	Length	418;		
Matches	144;	Conservative	68;	Mismatches	174;	Indels	74;	Gaps	15;		
Qy	53	VIGIALLALAIAGLGHFDCSGKY-RCRSEFCIELARCDGSPOCKDGEDEY-RC--	107	Qy	51	IVIGIALLALAIAGLGHFDC--SGYRCSSFKC--IELJARCGVSDOCKDGEDEY	106	Qy	51	IVIGIALLALAIAGLGHFDC--SGYRCSSFKC--IELJARCGVSDOCKDGEDEY	106
Db	872	VVGWVCKTPVMS-----CPODWLCHASEEPIVFLCDNVRACAGSDDESPHCKA	923	Db	22	IVWAGVW--ILAVTIALLVFLAQDKSYFYRSFQLINVEYNQLN---SPATOEW	74	Db	22	IVWAGVW--ILAVTIALLVFLAQDKSYFYRSFQLINVEYNQLN---SPATOEW	74
Qy	108	--VR-VGG--QNAVILQVFTAASWKTMCSDWKGHYANVACAGLQFPSSYSSDNLRVSS	160	Qy	107	CYRGGGONAVLQVFTAASWKTMCSDWKGHYANVACAGLQFPSSYSSDNLRVSS	157	Qy	107	CYRGGGONAVLQVFTAASWKTMCSDWKGHYANVACAGLQFPSSYSSDNLRVSS	157
Db	924	PLAYLVLVAGPTDREGVEINVHGTWGVCDDFGVREARVICRQLQFNGTAEVKSVYPP	983	Qy	161	LEQGFREEFVSDHLLP--DDKVTAHLHSVYREGASGHVTLQC-----	204	Db	75	TISGRISSLITKPFKEENLRNQFIRAHVAKLURGDSYVRAFDYVMKFTFTRNNINGASMKSR	134
Qy	205	-----TAGC-----HRYGSSRTVGGMNSLUSQWPWQSLOFQ	237	Db	1040	ATRPNPREFDVERSRKIHDPDAGRVLDPTLPTGAVYVPTKARPRLR	1039	Qy	161	LEQGFREEFVSDHLLP--DDKVTAHLHSVYREGASGHVTLQC-----	204
Db	984	GVGQIWIQDQVACNGTEPSIEDCV-----HWHWGEHNCATAEDVGVRGGVYPTKARPRLR	1039	Qy	238	GYHLCGGSVITPLWILITAHCYDLYLPKS-WTIQVG--LVSLLDNPAPSHLVEK-WVHS	294	Db	1100	TYHWCAGVLITRYVHLTAHCL--IGPKSTYVRIGDHYTAAYDNAEFLDIFIENYIHE	1157
Qy	205	-----TAGC-----HRYGSSRTVGGMNSLUSQWPWQSLOFQ	237	Qy	295	KYRP-KRUGNDIALMKLQPLTNEFQVCPUNSEENFPCKVCMVNGMATEDAGDA	353	Qy	214	SSRIVGGNMNLSLSOPWPHASLQFOGYHLCGGSVITPLWILITAHCYDLYLPKS-WTIQVG	273
Db	1040	ATRPNPREFDVERSRKIHDPDAGRVLDPTLPTGAVYVPTKARPRLR	1039	Db	1158	QFREGHMSNDIAVWVLLKTPVRFNDYVQPICLPARDOPALPQONCTISGWGATEAGSKDS	1217	Db	184	EORILGGEAEGSWPQVOSLRLNNNAHCCGSLINNNWILTAHCFRSNSNPRDWLATS	243
Qy	238	GYHLCGGSVITPLWILITAHCYDLYLPKS-WTIQVG--LVSLLDNPAPSHLVEK-WVHS	294	Qy	354	SYVLNHAAPLNSNKTCNRDVYGGITSPSMICAGYLGVDSQEDSGGLPLVUC-QERRL	412	Db	135	IESVLRQMLNNNSGNLE-INPSREITSLTDQAAANLWINECGAGPDLTLS-----	183
Db	1100	TYHWCAGVLITRYVHLTAHCL--IGPKSTYVRIGDHYTAAYDNAEFLDIFIENYIHE	1157	Db	1218	STDLRACTVPLPDSVCRPEPGLSDGMSDQSGGLPLVCPNSEGL	1277	Db	244	ISTTF--PKLRKRYVNLINQHNNYSATHENDALVRYLNYSVTFKDHISVCPAATQON	301
Qy	295	KYRP-KRUGNDIALMKLQPLTNEFQVCPUNSEENFPCKVCMVNGMATEDAGDA	353	Qy	334	DGKVCWTSGGATEDGADASPVNLNHAAPLTSNKICNRDVYGGITSPSMICAGYLGVDSQEDSGGLPLVUC-QERRL	393	Db	302	POSTAVYTGWA--QEVAGHTVYELRQQVRITISNDVNCNAHPSYNGAISLGMICAGVPGQG	360
Qy	413	WKLNGATSGFGICCAEVNKGPGVYTRVTSFLDMHIEQERDL	452	Qy	394	VISCOQGSGGPLVUCQEE-RRLWKLNGATSGFGICCAEVNKGPGVYTRVTSFLDMHIEQ	447	Db	361	VDACQGDSSGGPLVQEDSRRLWFLIVGIVSWGDQGLPDKPGVYTRVYALDWIRQ	415
Db	1218	STDLRACTVPLPDSVCRPEPGLSDGMSDQSGGLPLVCPNSEGL	1277	Db	1278	HTLTGTVSWGKHCGYANKPGVYVLYKVAHYRDWIEQKLNQL	1317	Db	361	VDACQGDSSGGPLVQEDSRRLWFLIVGIVSWGDQGLPDKPGVYTRVYALDWIRQ	415
RESULT	8			RESULT	9			RESULT	9		
QNATO				QNATO				QNATO			
ID	060235	PRELIMINARY;	PRT;	ID	060235	PRELIMINARY;	PRT;	ID	060235		
AC	060235;			AC	060235;			AC	060235;		
DT	01-AUG-1998	(TREMBlre. 07, Created)		DT	01-AUG-1998	(TREMBlre. 07, Last sequence update)		DT	01-AUG-1998	(TREMBlre. 07, Last sequence update)	
DT	01-JUN-2001	(TREMBlre. 17, Last annotation update)		DT	01-JUN-2001	(TREMBlre. 17, Last annotation update)		DT	01-JUN-2001	(TREMBlre. 17, Last annotation update)	
DE		AIRWAY TRYPSIN-LIKE PROTEASE.		DE		ADHESIVE SERINE PROTEASE.		DE		SP222D.	
OS		Homo sapiens (Human).		OS		Anopheles gambiae (African Malaria mosquito).		OS		Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		OC		Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;		OC		Anopheles.	
OX		NCBI_TaxID:9606;		OX		NCBI_TaxID:7165;		OX			
RN		[1]		RN		[1]		RN			
RP		SEQUENCE FROM N.A.		RP		SEQUENCE FROM N.A.		RP			
RX		MEDLINE#98234382; PubMed=9565616;		RX		MEDLINE#20318993; PubMed=10860981;		RX			
RA		Yamaoka K., Masuda K., Ogawa H., Takagi K., Umemoto N., Yasuoka S.:		RA		Danieli A., Loukeris T., Lagueux M., Mueller H.M., Richman A.,		RA			
RT		"Cloning and characterization of the cDNA for human airway trypsin-like protease.;"		RT		Kafatos F.C.;		RT			
RL		J. Biol. Chem. 273:11895-11901(1998).		RL		"A modular chitin-binding protease associated with hemocytes and hemolymph in the mosquito Anopheles gambiae;"		RL			
CC		- SIMILARITY: TO SERINE PROTEASES TRYPSIN FAMILY.		CC		- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A		CC			
DR		- SIMILARITY: TO CHMOTRIPSIN SERINE PROTEASE FAMILY (S1).		DR		(LDLR) DOMAIN.		CC			
EMBL		AB002134; BA28691.1; -		EMBL		- SIMILARITY: TO SERINE PROTEASES TRYPSIN FAMILY.		CC			
DR		HSSP: P0750; I-REP.		DR		- SIMILARITY: TO CHMOTRIPSIN SERINE PROTEASE FAMILY (S1).		CC			
DR		MEROPS; S01.301; -		DR		- SIMILARITY: TO CHMOTRIPSIN SERINE PROTEASE FAMILY (S1).		CC			
DR		InterPro; IPR001314; Chymotrypsin.		DR		EMBL: AU276428; CAB81934.1; Chitin_bind_2.		CC			
DR		InterPro; IPR000082; SEA.		DR		InterPro; IPR02557; Chitin_bind_2.		CC			
DR		InterPro; IPR001254; Trypsin.		DR		InterPro; IPR01314; Chymotrypsin.		CC			
DR		Pfam; PF00189; SEA; 1.		DR		InterPro; IPR02172; LDL_recept_A.		CC			
DR		PRINTS; PRO0089; Trypsin; 1.		DR		InterPro; IPR001190; SRCR.		CC			
DR		SMART; SM00200; Tryp_SPC; 1.		DR		InterPro; IPR01254; Trypsin.		CC			
DR		PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.		DR		Pfam; PF00057; ldl_recept_a; 2.		CC			
KW		Hydrolase; protease; serine protease.		DR		Pfam; PF00530; SRCR; 2.		CC			
SQ		SEQUENCE 418 AA; 46263 MW; F4BC1DB020CFBBDD CRG64;		DR		Pfam; PF00089; Trypsin; 1.		CC			
		PRINTS; PRO0261; LDLRECEPTOR.		DR		PRINTS; PRO0258; SPERACTRCPTR.		CC			

DE	SEARINE PROTEASE).
OS	<i>rattus norvegicus</i> (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; <i>Rattus</i> .
RN	NCBI_Taxid=10116;
RP	[1] SEQUENCE FROM N.A.
RC	SPRAIN-FISTAR; TISSUE=DUODENUM;
RA	TSUZUKI S.;
RT	"A membrane bound serine protease expressed in rat small intestine.";
RL	Submitted (JAN-2000) to the EMBL/Genbank/NCBI databases.
RP	[2] SEQUENCE FROM N.A.
RC	SPRAIN-FISTAR; TISSUE=DUODENUM;
RA	Inoue H., Takahashi K., Kishi K.;
RT	"Membrane-bound arginine specific serine protease.";
RL	Submitted (SEP-2000) to the EMBL/Genbank/NCBI databases.
CC	-1 SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC	-1 SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
CC	-1 SIMILARITY: TO CHYMOtrypsin SERINE PROTEASE FAMILY (S1).
DR	EMBL: AB037898; BAB03022.1. -.
DR	EMBL: AB049189; BAB13765.1. -.
DR	InterPro: IPR02106; AA_5'RNA_ligase_II.
DR	InterPro: IPR01314; Chymotrypsin.
DR	InterPro: IPR00059; CUB.
DR	InterPro: IPR00172; LDL_recept_A.
DR	InterPro: IPR01254; trypsin.
DR	Pfam: PF00431; CUB; 2.
DR	Pfam: PF00057; LDL_recept_A; 4.
DR	Pfam: PF00059; trypsin; 1.
DR	PRINTS: PRO0722; CHYMOtrypin.
DR	PRINTS: PRO0261; LDLRECEPTOR.
DR	SMART: SM00042; CUB; 2.
DR	SMART: SM00192; LDLA; 3.
DR	SMART: SM00020; tryp_SPC; 1.
DR	PROSITE: PS00339; AA_5'RNA_ligase_II_2; UNKNOWN_1.
DR	PROSITE: PS01180; CUB; 2.
DR	PROSITE: PS01209; LDLA_1; 2.
DR	PROSITE: SS0006; LDLA_2; 4.
DR	PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR	PROSITE: PS00135; TRYPSIN_SER; 1.
KN	GLYcoprotein; Hydrolase; Protease; Serine protease.
FT	VARIANT 665 665 K -> N.
SEQUENCE	855 AA; 9495 MW; 3580657ECF6CF03D CRC64;
Query	Match 22.2%; Score 542.5; DB 11; Length 855; Best Local Similarity 31.5%; Pred. No. 3e-42; Indels 83; Gaps 16; Matches 135; Conservative 63; Mismatches 148; Indels 83; Gaps 16;
QY	73 CSGKVRORSFKCIELIARCDGSVDSCKDGDEYRCVRYGQNAVLQVFTAASWKTMCSDD 132
Db	453 CGEMGEMKTCG-KCIRKDLRCDGWCADPDYSDERHC-----RCNATHQFCKNQ 499
QY	133 -----WKGHyanvacQQLGFPsyVSDNLRVSSLQGREFVSDHLPD----- 178
Db	500 FCKPLFWCDSN-DCGD-----GSDEEGCSCPAGSKF--CSNQKCLFOSQOCNGKD 548
QY	179 -----DKVTAhHsVYvreGqASHVtLoCTACahRRGTS----- 214
Db	549 DGGDGSDEASCNDVNAYVSCTKYTYR-CQNGICLNGKNPEDCOKKDCSDGSEBKNCDCGLR 607
QY	215 -----SRVGGNMSLSSQWPMQASLQFQGY-HLGGSVITPLWITAACHCYDLYPK- 266
Db	608 SITKQARVGGNNADEEWPMQVSHLQHGQHGLGASLSPWVLVRAHCQDFTFKYS 667
RL	Science 267-2185-2195(2000).
CC	-1 SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A
CC	(LDL) DOMAIN.
CC	-1 SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC	-1 SIMILARITY: TO CHYMOtrypsin SERINE PROTEASE FAMILY (S1).
DR	EMBL: AAF0553; AAF0519.2. -.
RESULT	12
Q9VSU2	
ID	Q9VSU2 PRELIMINARY; PRT; 1186 AA.
AC	Q9VSU2; Q9VSU1;
DT	01-MAY-2000 (Tremblel. 13, Created)
DT	01-MAR-2001 (Tremblel. 16, Last sequence update)
DT	01-JUN-2001 (Tremblel. 17, Last annotation update)
DE	TEQUILA PROTEIN.
GN	TEQUILA OR CG4821 OR CG18403.
OS	Drosophila melanogaster (Fruit fly).
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.
OC	NCBI_Taxid=7227;
RN	[1] SEQUENCE FROM N.A.
RC	SPRAIN-BERKELEY; MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers J.-H.C., Blazquez R.G., Champé M., Pfeiffer B.D., Wan K.-H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abil J.F., Agbayani A., An H.-J., Andrews Pfannkoch C., Baldwin D., Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Centner A., Chandra I., Cherry J.M., Cawley S., Dahake C., Davenport L.B., Davies P., de Pablo J., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Doobson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fosler C., Gabriele A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Houston K.A., Houston K.A., Howland T.J., Wei M.-H., Hsiegwan T.C., Jalali M., Houston K.A., Hsien F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Leis J., Levittsky A., Li J., Lill J., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milashina N.V., Mobarry C., Morris J., Morris J., Mosherfi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson J., Nixon K., Nuskeen D.R., Pacble J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svartkas R., Tector C., Turner R., Venter J., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Yeh J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.-H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C., RT "The genome sequence of <i>Drosophila melanogaster</i> ."
Db	RPLCLPDNTVHFVPGAKATWWTGNGHTKG-GTCGALLQKGTEVINGQTC-EELPQQI 784
QY	381 SP8NLICAGLTYGGVDSQCGDGGPLCQER -RLWKLVQATFGIGAEVNRKGVYTRV 438
Db	765 TPRMMCVFLLGGVDSQCGDGGPLSSVEKGIFQ-RGVVNGEGRQRNKGPGVYRIP 843
QY	439 SELDWIHEQ 447
Db	844 EVRDWIKEQ 852

DR HSSP; P00231; 1AQO.
 DR FBgn023479; Tequila.
 DR InterPro; IPR00194; Kmpase_alpha_beta.
 DR InterPro; IPR00134; Chymotrypsin.
 DR InterPro; IPR002172; LDI_recept_A.
 DR InterPro; IPR00295; P_rich_extens.
 DR InterPro; IPR001190; SRCR.
 DR InterPro; IPR001234; Trypsin.
 DR Pfam; PF00057; Idl_recept_a; 2.
 DR Pfam; PF00530; SRCR; 2.
 DR Pfam; PF00089; trypsin; 2.
 DR PRINTS; PR00722; CHYMO TRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR PRINTS; PR01217; PRICHEXTNS.
 DR PRINTS; PR00258; SPERACTRCPR.
 DR SMART; SM00192; LDLa; 2.
 DR SMART; SM00202; SRCR; 2.
 DR SMART; SM00220; TRYp_SPEC; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS01209; LDLA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_2.
 DR PROSITE; PS0287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR KW Glycoprotein; Hydrolase; Serine protease.
 SQ SEQUENCE 1186 AA; 131246 MW; F3474D3968B8A32D CRC64;

Query Match 22.0%; Score 539; DB 5; Length 1186;
 Best Local Similarity 27.5%; Pred. No. 9.9e-42; Mismatches 134; Conservative 74; Indels 167; Gaps 15; Matches 133; Chaitin.bind; 2; DR P00152; ATPASE_ALPHA_BETA; UNKNOWN_1.

QY - 66 GLGIH-----FDCSGKY-RCKSSFKCIELIARCDGVSDCKGEDEYRCV 108
 Db 709 GMGVHNCVGDDEVAGVTCKVPKVMCPNNYWLCHTSKECIPPAVCDNTPDCADSKSDE--CA 766

QY 109 RV-----GGONA---VIQVFTAAWSKTMCSDDWKGHYANVACAOQLGPPSYSSDN 155
 Db 767 AVQCAPVQYRLEGRGRNSNEGRLEVKHHGKVWGSVCDDFNLKSAQVACNSMGFFGPAKIE 826

QY 156 LRVSSLCQFREEFVSDHLL--PDKKVTAHHHSVYVREGGASGHVTLOCTA----- 206

Db 827 NFGNSNQP----IWLDQVMCFQNETSIDQCNHWNNGEHNCHNTEDVALHC SAGPPRS 881

QY 207 -----CG-----HRGYS 214

Db 882 QRYSQTQIKGGRSLGRETPKTKYQIGIWERSSKAVITPREGIFKDLTDIYAHR--E 938

QY 215 SRIVGGNMSSLSQWPWQASLQFQG----YHLCGGSVITPLWITAACVYDYLPLK-SW 268

Db 939 ERVVRGNVAQQRGRHWPQATIRTRGRRGGI-SPKGAY 996

QY 269 TQVQ--LVSLLDNPAHSLVKEKWHISKY-KPKRLGNDIALMKGALQPLFEMIQVCL 325

Db 997 FVFRGVHDYNTAASSEYDSTFIENWYLHENFRKGTHMNDIALWVLPKTFESDDYVOPICL 1056

QY 326 PNSSEENFDGKVWTSQHGATEDAGDASPVLINHAAPLISNKICNRDVGGIIPSML 385

Db 1057 PPKNAELVEDRKCTISGWSKISGSVSTPAQVQLGSAELPILADHVKCOSNVNGSSMSGMF 1116

QY 386 CAGYLTGGVDSQCGDSSGQPLVQOERRLWKLVQTSATSGFSGICAEVNPKPSVYTRVTSFLWTH 445

Db 1117 CAGSMDESVDACEGQDGGPLVCSDDGGETLYGLISWMQHCFKKNRPGVYVVRVHYIDWIY 1176

QY 446 EUDERDL 452

Db 1177 EKINESL 1183

Query Match 22.0%; Score 539; DB 5; Length 1449;
 Best Local Similarity 27.5%; Pred. No. 1.3e-41; Mismatches 134; Conservative 74; Indels 167; Gaps 15; Matches 133; Chaitin.bind; 2; DR P00152; ATPASE_ALPHA_BETA; UNKNOWN_1.

QY 66 GLGIH-----FDCSGKY-RCKSSFKCIELIARCDGVSDCKGEDEYRCV 108
 Db 972 GMGVHNCVGDDEVAGVTCKVPKVMCPNNYWLCHTSKECIPPAVCDNTPDCADSKSDE--CA 1029

QY 109 RV-----GGONA---VIQVFTAAWSKTMCSDDWKGHYANVACAOQLGPPSYSSDN 155
 Db 1030 AVQCAPVQYRLEGRGRNSNEGRLEVKHHGKVWGSVCDDFNLKSAQVACNSMGFFGPAKIE 1089

QY 156 LRVSSLCQFREEFVSDHLL--PDKKVTAHHHSVYVREGGASGHVTLOCTA----- 206

Db 1090 NFGNSNQP----IWLDQVMCFQNETSIDQCNHWNNGEHNCHNTEDVALHC SAGPPRS 1144

QY 207 -----CG-----HRGYS 214

Db 1145 QRYSQTQIKGGRSLGRETPKTKYQIGIWERSSKAVITPREGIFKDLTDIYAHR--E 1201

QY 215 SRIVGGNMSSLSQWPWQASLQFQG----YHLCGGSVITPLWITAACVYDYLPLK-SW 268

AC Q9UI12; DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DR 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DR 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DR GRAL PROTEIN
 DR TEQUILA OR GRAL OR CG4821 OR CG4948 OR CG18403.
 DR OS Drosophila melanogaster (Fruit fly).
 DR OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 DR OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 DR OC Ephydriidea; Drosophilidae; Drosophila.
 DR NCBI_TAXID=7227;
 DR RN [1];
 DR RP SEQUENCE FROM N.A.
 DR RA Munier A.-I., Medzhitov R., Janeway C., Hoffmann J.A., Lagueux M.;
 DR RT "Characterization of a new serine protease in Drosophila melanogaster";
 DR RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR CC -1 - SIMILARITY: TO LOW DENSITY LIPROTEIN (LDL) RECEPTOR CLASS A
 DR CC (LDLRA) DOMAIN.
 DR CC -1 - SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR CC -1 - SIMILARITY: TO CHYMO TRYPSIN SERINE PROTEASE FAMILY (S1).
 DR EMBL; A251803; CAB6453.1; -
 DR HSSP; P20231; 1AQO.
 DR FBgn023479; Tequila.
 DR InterPro; IPR00194; ATPASE_alpha_beta.
 DR InterPro; IPR02557; Chitin_binding.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR02172; LDL_recept_A.
 DR InterPro; IPR01190; SRCR.
 DR InterPro; IPR01254; Trypsin.
 DR Pfam; PF001607; Chitin_bind; 2.
 DR Pfam; PF00537; Idl_recept_a; 2.
 DR Pfam; PF000530; SRCR; 2.
 DR PRINTS; PR00089; trypsin; 1.
 DR SMART; SM00202; CHYMO TRYPSIN.
 DR PRINTS; PR00722; CHYMO TRYPSIN.
 DR SMART; SM00192; LDLa; 2.
 DR SMART; SM00202; SRCR; 2.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS01209; LDLRA_1; 1.
 DR PROSITE; PS50068; LDLRA_2; 2.
 DR PROSITE; PS00420; SRCR_1; UNKNOWN_2.
 DR PROSITE; PS0287; SRCR_2; 2.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00115; TRYPSIN_SER; 1.
 DR KW Glycoprotein; Hydrolase; Serine protease.
 SQ SEQUENCE 1449 AA; 16045 MW; F35C806543CAB6ED CRC64;

QY	269	TIOVG--LVSLLDPAPSHLVEKIVHSKY-KPRKLGNIDALMLKLAGLTFNEMIQPVCL	325	1202	ERVVRGNGVAQRGRPHQWQATIRTRGCGTSSHQGAVVSKRHLITAACRHLYG--SPKGAY	1259	
QY	1260	FVRVGDHYANIAESSEVDTSFIENWYLHENFRKGTHMNNIDALVLUKTPKFSVDYOPICL	1319	Db	1320	PDKNAELVEDRKCTISGWSIKSGVSVPAQVLGSAELPLADIVCKOSNVYGSAMSEGMF	1379
QY	386	CAGYLTGGVDSQCGDGGPLVQERRLWKLVGAATSGFGCAEVNPKGVYTRYTFLWHL	445	Db	1380	CAGSMDESVDACEGDSGGPLVCSDDDGGETLYGLISW5QHCGFKNRPGVYVRVNHYIDWY	1439
QY	446	EQMERDL	452	Db	1440	EKINSEL	1446
RESULT	14						
Q9UI13		PRELIMINARY;	PRT;	1462 AA.			
ID	Q9UI13						
AC	Q9UI13;						
DT	01-MAY-2000	(TREMBREL. 13, Created)					
DT	01-MAY-2000	(TREMBREL. 13, Last sequence update)					
DT	01-JUN-2001	(TREMBREL. 17, Last annotation update)					
DE	GRAM PROTEIN.						
GN	TEQUILA OR GRAL1 OR CG4821 OR CG4948 OR CG18403.						
OS	Drosophila melanogaster						
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Drosophilidae; Drosophila.						
OC	Phylogroup: Drosophila						
OC	OX						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	Munier A.I., Medzhitov R., Janeway C., Hoffmann J.A., Lagueux M.;						
RT	"Characterization of a new serine protease in Drosophila melanogaster."						
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.						
CC	-!- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A (LDLRA) DOMAIN.						
CC	-!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.						
CC	-!- SIMILARITY: TO CHYMOTRYPsin SERINE PROTEASE FAMILY (S1).						
DR	EMBL; AJ251802; CAB64652.1; -.						
DR	RSPB; P00750; IRTF.						
DR	FlyBase; FBgn0023479; Tequila.						
DR	InterPro; IPR00194; ATPase_alpha_beta.						
DR	InterPro; IPR002557; Chitin_binding.						
DR	InterPro; IPR001314; Chymotrypsin.						
DR	InterPro; IPR002172; LDL_recept_A.						
DR	InterPro; IPR002965; P_rich_extensn.						
DR	InterPro; IPR001190; SRCR.						
DR	InterPro; IPR001254; Trypsin.						
DR	Pfam; PF001607; Chitin_bind_2; 2.						
DR	Pfam; PF00057; ldl_recept_a; 2.						
DR	Pfam; PF00520; SRCR; 2.						
DR	Pfam; PF00049; trypsin; 1.						
DR	PRINTS; PR0072; CHYMOTRYPSIN.						
DR	PRINTS; PR00261; LDLRECEPTOR.						
DR	PRINTS; PR01217; PRICHEXTENSN.						
DR	PRINTS; PR00258; SPERACTRCPTR.						
DR	SMART; SM00494; ChtBD2; 2.						
DR	SMART; SM00192; LDLra; 2.						
DR	SMART; SM00202; SR; 2.						
DR	SMART; SM00200; TRYSPC; 1.						
DR	PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.						
DR	PROSITE; PS01209; LDLRA_1; 1.						
DR	PROSITE; PS00066; LDLRA_2; 2.						
DR	PROSITE; PS00420; SRCR_1; UNKNOWN_2.						
DR	PROSITE; PS02027; SRCR; 2.						
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.						
DR	PROSITE; PS00135; TRYPSIN_SER; 1.						
RESULT	14						
Q9UI13		PRELIMINARY;	PRT;	1462 AA.			
ID	Q9UI13						
AC	Q9UI13;						
DT	01-MAY-2000	(TREMBREL. 13, Created)					
DT	01-JUN-2001	(TREMBREL. 13, Last sequence update)					
DT	01-JUN-2001	(TREMBREL. 17, Last annotation update)					
DE	GRAM PROTEIN.						
GN	TEQUILA OR GRAL1 OR CG4821 OR CG4948 OR CG18403.						
OS	Drosophila melanogaster (Fruit fly).						
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Drosophilidae; Drosophila.						
OC	Phylogroup: Drosophila						
OC	OX						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	Munier A.I., Medzhitov R., Janeway C.A., Lanot R., Zachary D., Capovilla M., Lagueux M.;						
RA	Capovilla M., Lagueux M.;						
RT	"Gral1 a Drosophila gene coding for several mosaic serine proteases."						
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.						
DR	EMBL; AU309005; CAC35209.1; -.						
QY	207						
Db	1158 QRYSQTQIKGGRSLGRETTPKTYSQIGLWERSKAVHPTPRRCGFKDDLTDEYAH--E	1214					
QY	215 SRYVGGNMSLILSONQWQASLQFOG----YHLCGGSVITPLWITTAHCVYDLYLPK-SW	268					
Db	1215 ERVVGNGVAQRGRHFWQATIRTRGFRGGISSHWCGAVVSKRHLITAACRHLYG--SPKGAY	1272					
QY	269 TIOVG--LVSLLDPAPSHLVEKIVHSKY-KPRKLGNIDALMLKLAGLTFNEMIQPVCL	325					
Db	1273 FVRVGDHYANIAESSEVDTSFIENWYLHENFRKGTHMNNIDALVLUKTPKFSVDYOPICL	1332					
QY	326 PNSEENFPDKVCTISGWMGATEDGADASPVLNHAAPLINSKICCNHRDVGGITPSML	385					
Db	1333 PDKNAELVEDRKCTISGWSIKSGVSVPAQVLGSAELPLADIVCKQSNVYGSAMSEGMF	1392					
QY	386 CAGYLTGGVDSQCGDGGPLVQERRLWKLVGAATSGFGCAEVNPKGVYTRYTFLWHL	445					
Db	1393 CAGSMDESVDACEGDSGGPLVCSDDDGGETLYGLISW5QHCGFKNRPGVYVRVNHYIDWY	1452					
QY	446 EQMERDL	452					
Db	1453 EKINSEL	1459					
RESULT	15						
Q9B119		PRELIMINARY;	PRT;	2382 AA.			
ID	Q9B119						
AC	Q9B119;						
DT	01-JUN-2001	(TREMBREL. 17, Created)					
DT	01-JUN-2001	(TREMBREL. 17, Last sequence update)					
DT	01-JUN-2001	(TREMBREL. 17, Last annotation update)					
DE	GRAL2 PROTEIN PRECURSOR.						
GN	GRAL1						
OS	Drosophila melanogaster (Fruit fly).						
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Drosophilidae; Drosophila.						
OC	Phylogroup: Drosophila						
OC	OX						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	Munier A.I., Medzhitov R., Janeway C.A., Lanot R., Zachary D., Capovilla M., Lagueux M.;						
RA	Capovilla M., Lagueux M.;						
RT	"Gral1 a Drosophila gene coding for several mosaic serine proteases."						
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.						
DR	EMBL; AU309005; CAC35209.1; -.						
QY	2382						
FT	SIGNAL	1	23	POTENTIAL			
SQ	SEQUENCE	2382 AA;	264348 MW;	51C85282B0683D4 CRC64;			
Query	Match	22.0%	Score	539; DB 5; Length	2382;		
Best	Local	Similarity	27.5%	Pred. No.	2.5e41;		

Matches	Conservative	74;	Mismatches	167;	Indels	112;	Gaps	15;						
Qy	66	GLCLIH-----	-----EFC	SGKY-RCR	ESFKCIELIARCDG	YSDDMKHYANVACQ	LGPPSYVSSDN	108						
Db	1905	GWGVHNCGVDEAVGVTCKV	PVMKCPNNWLCHSK	ECI	PPAFVCDNT	PDCA	DRSDE-CA	1962						
Qy	109	RV-----GGONA--VLO	UFTAASW	KINCSDDMKHYANVACQ	LGPPSYVSSDN	155								
Db	1963	AVQCAPVYRLEGGRNSNEGR	EVKHHG	WSVCDDEFNLKSA	QVACNSMGFP	PAKIEK	2022							
Qy	156	LRRSSLEGOFREEFVSI	DHLL--PDKV	TALHHSVYREGC	ASGHV	VTLQTA-----	206							
Db	2023	NTIGNSNCP-----IWLDQ	WMC	QNETSIDQCNH	WNGEHN	CHTEDVALICSSAGPPRS	2077							
Qy	207	-----	-----	-----	-----	-----								
Db	2078	QRTSQTIKGGRSLGRE	TPKTYSQIGL	WERSSSKAVH	TPRCC	IFKDLTDEYAH	HRGYS 214							
Qy	2115	SRVGGNNM	LLS	WPO	SLORG-----YHIC	GGGSVITPLW	ITAAMCVYDYL	LPK-SW 268						
Db	2135	ERVVRG	YNAQRGR	HPWQAT	RTKG	GGGTS	SHWMGAVV	SKRHLTAAMCLY	G--SPKGAY 2192					
Qy	269	TIQVG--LVS	LDNPAPSHL	VEIVHISKY	-KPKR	LGNDIALM	KLAGPLTF	NEMIQVCL 325						
Db	2193	FVRVGDH	YANIA	BSSEVDFIEN	WYHLEN	FRKGTH	MNDIALV	VTKPLKFSDYVQ	PICL 2252					
Qy	326	PNSBENFPDGK	VWTS	SGM	GATEDGAGD	ASPVLINHA	AVPLISN	ICNHRDVYGGTIS	PSML 385					
Db	2253	PDRNAELV	EDRKCT	TSIG	WGS	IKSGV	STPAQVL	GSALPILAD	WCKQSNVYGSMSBGMF 2312					
Qy	386	CAYLTG	GVDS	CGD	SGSPLV	CCER	RLKLVG	TSGFGC	AEVNPKG	YTRYV	TSFLW	W 445		
Db	2313	CAGSM	DES	DV	ACG	PLV	CDDG	DTLY	GLJIS	WGOH	CGFKN	RPGVY	VRYV	WYIDWY 2372
Qy	446	EQERDL	452	1:	1:	1:	1:	1:	1:	1:	1:	1:	1:	1:
Db	2373	EKINESL	2379	1:	1:	1:	1:	1:	1:	1:	1:	1:	1:	1:

Search completed: April 11, 2002, 08:54:33
 Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 08:52:32 ; Search time 19.48 Seconds

Sequence: 1 MGENDPPAVEAPFSRSLFG.....TRVTSFLDWIHEQMRDLKT 454
(without alignments)
1775.321 Million cell updates/sec

Title: US-09-846-512-12
Perfect score: 2447
Sequence: 1 MGENDPPAVEAPFSRSLFG.....TRVTSFLDWIHEQMRDLKT 454
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS

Result No.	Score	Query	Match Length	DB ID	Description	RESULT	1
1	664.5	27.2	417	1	500845 hepsin (EC 3.4.21.-) - human	500845	
2	660	27.0	1035	1	A43090 enteropeptidase (E	C;Species: Homo sapiens (man)	
3	657	26.8	1019	1	A45318 enteropeptidase (E	C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999	
4	654	26.7	1034	1	A45363 enteropeptidase (E	C;Accession: 500845	
5	645	26.4	416	1	S33777 hepsin (EC 3.4.21.-)	R;Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.	
6	570	23.3	638	1	KORTPL plasma kallikrein	Biochemistry 27, 1067-1074, 1988	
7	561	22.9	638	1	KQHUP plasma kallikrein	A;Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane	
8	540	22.1	638	1	PLPG plasma kallikrein	A;Reference number: 500845; MID:88209431	
9	526	21.5	761	2	JC5759 brain-specific ser	A;Accession: 500845	
10	523.5	21.4	523.5	21.4	A61545 plasmin (EC 3.4.21.-)	A;Molecule type: mRNA	
11	520.5	21.3	455	2	A61545 plasmin (EC 3.4.21.-)	A;Residues: 1417 <LEY>	
12	507.5	20.7	812	1	PLMS plasmin (EC 3.4.21.-)	A;Cross-references: EMBL:X07732; NID:932063; PIDN:CAA30558.1.; PID:932064	
13	507	20.7	437	2	S18407 acrosin (EC 3.4.21.-)	C;Genetics:	
14	506	20.7	1113	2	JDE0315 low-density lipopr	A;Cross-references: GDB:335685; OMM:142440	
15	504.5	20.6	810	1	PLHU plasmin (EC 3.4.21.-)	C;Superfamily: hepsin; trypsin homology	
16	503.5	20.6	273	2	A47246 tryphase (EC 3.4.21.-)	C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein	
17	503.5	20.6	436	2	JX0172 acrosin (EC 3.4.21.-)	F;153-400/Domain: trypsin homology <TR>	
18	503	20.6	460	1	B61545 plasmin (EC 3.4.21.-)	F;180-204,291-359,322-338,349-381/Disulfide bonds: #status predicted	
19	501.5	20.5	810	2	146260 plasmin (EC 3.4.21.-)	F;203,257-353/Active site: His, Asp, Ser #status predicted	
20	500.5	20.5	275	2	A32410 tryphase (EC 3.4.21.-)	Query Match, Best Local Similarity 35.7%; Pred. No. 1.5e-45; DB 1; Length 417; Matches 149; Conservative 62; Mismatches 143; Indels 63; Gaps 12;	
21	500	20.5	418	1	A37344 acrosin (EC 3.4.21.-)	Db 59 LILAAIGIGIHFQDSKGKVRCRSSPKCIELIARDGVSKDGDDEYRVRVQGQAVLQ 118 Db 26 LILLTAIG-----AASPAIVAVLRLR-----QEPLYVQVSQDARM 64	
22	500.5	20.5	421	1	S11674 tryphase (EC 3.4.21.-)	Db 177 PDDKTYALHHSVYVRE----GCASGHVWVLTQFACGHRRGYSRISVGNNMSLISOWHQ 231 Db 123 -----GRLPHTQRLIEVIVSVCDCPGRGRFLAICQDCGRRKLPVDRIVGGRRDTSLGWRPQW 177	
23	498.5	20.4	431	2	S47538 acrosin (EC 3.4.21.-)	Db 178 VSLRDGAHCGGSILSGWVLTAAHC----FERNRVLISRWFAGAWA---QASPIG 229	
24	498	20.4	460	1	PLOB plasmin (EC 3.4.21.-)	Db 123 -----GRLPHTQRLIEVIVSVCDCPGRGRFLAICQDCGRRKLPVDRIVGGRRDTSLGWRPQW 177	
25	497.5	20.3	810	2	B30848 plasmin (EC 3.4.21.-)	Db 230 LQIGQAVWVYHGGLPFRPNESEINSNDALVHLSPLTEVQPCVLPAAQALVQK 289	
26	495.5	20.2	270	2	S55160 mast cell tryptase	Db 337 VCWTSGWGATEDAGDASPVLNHAAPLISNKICNHRDYYGIGITSPSMICAGYLGGDS 396	
27	494	20.2	625	1	KFHU1 coagulation factor		
28	490.5	20.0	276	2	A38654 mast cell proteina		
29	488.5	20.0	274	2	JC4171 tryptase (EC 3.4.21.-)		

Db	290	ICTVIGWGWNTQ-YVGQDAGSVLQEARVPTISNDYVNGADPYGQNLKPKMCAGYEGGIDDA	348	QY	397	CQGDGGGPVNCQE---RRLWKLYGATEFGIGCAEVNKPGVYTRVTSFLDWIHRQME	449	QY	349	CQGDGGPVCEDSISRPRWRLCGIVSGKGTGCAQKPGVYTRVSDFREWIQFAIK	405
RESULT	2										
A43090		enteropeptidase (EC 3.4.21.9) precursor - bovine									
N;Alternate names: enterokinase											
C;Species: Bos primigenius taurus (cattle)											
C;Accession: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999											
R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.											
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994											
A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of intestinal enteropeptidase and trypsin.											
A;Reference number: A43090; MUID:94329561											
A;Accession: A43090											
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ											
A;Molecule type: mRNA											
A;Cross-references: GB:U09859; PIDN:9746410; PIDN:AAB40026.1.; PID:9746411											
A;Experimental source: small intestine											
R;Levalle, E.R.; Rehentulla, A.; Racine, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.; J. Biol. Chem. 268, 23311-23317, 1993											
A;Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase.											
A;Reference number: A48874; MUID:94043122											
A;Accession: A48874											
A;Molecule type: mRNA											
A;Residues: 801-1035 <LAV>											
A;Cross-references: GB:U119663; PIDN:9416131; PIDN:AAA16035.1; PID:9416132											
A;Note: parts of this sequence, including the amino end of the mature protein, were confirmed by Edman sequencing.											
R;Light, A.; Jansta, H.											
J. Protein Chem. 10, 475-480, 1991											
A;Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.											
A;Reference number: A61436; MUID:92189715											
A;Accession: A61436											
A;Molecule type: protein											
A;Residues: 801-807, 'Y', 809-827 <LIG>											
C;Comment: The mechanism of association with the membrane of the intestinal brush border membrane attachment using a signal-anchor sequence.											
C;Comment: Conversion from membrane-bound to soluble forms may involve further processing.											
C;Complex: mature enteropeptidase is variously reported to contain two (heavy and light) isoflde linked											
C;Function:											
A;Description: cleaves propeptide from trypsinogen to produce active trypsin											
A;Pathway: intestinal digestive hydrolytic cascade											
A;Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding repeat homology											
C;Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein											
F;22-38/domain: transmembrane #status predicted <TMM>											
F;52-117/product: enteropeptidase mini chain #status predicted <MCH>											
F;118-800/product: enteropeptidase heavy chain #status predicted <RCH>											
F;199-236/domain: LDL receptor ligand-binding repeat homology <LDL1>											
F;358-520/domain: MAM homology <MAM>											
F;542-647/domain: C1r/C1s repeat homology <C1R>											
F;659-683/domain: LDL receptor ligand-binding repeat homology <LDL2>											
F;694-799/domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC>											
F;801-1035/product: enteropeptidase light chain #status predicted <CLH>											
F;116-147,170,194,233,263,264,404,456,519,550,646,698,722,741,762,864,903,965/Bindin											
F;788-912,826-842,926-993,957-972,981-1011/Disulfide bonds: #status predicted											
F;841,892,987/Active site: His, Asp, Ser #status predicted											
Query Match 27.0%; Score 660; DB 1; Length 1035;											
Best Local Similarity 36.9%; Pred. No. 1e-44;											
Matches 146; conservative 70; Mismatches 146; Indels 34; Gaps 15;											
Db	290	ICTVIGWGWNTQ-YVGQDAGSVLQEARVPTISNDYVNGADPYGQNLKPKMCAGYEGGIDDA	348	QY	121	TAASMTKMCSDWKHYANVACAQDGFPSVSSDNRVSLQEGOREEYFISDHLLPDDK	180	QY	121	TAASMTKMCSDWKHYANVACAQDGFPSVSSDNRVSLQEGOREEYFISDHLLPDDK	180
Db	711	IOTSHVCAENWTQISDDVQCOLLGLT--GNSSVPTFSTGG--GPYVNLL-----	758	QY	181	WTALHHSVY--RECSAGHVLQVLTCT--ACGHR--RGSSRIVGNNMSLQSPWQAS	233	QY	181	WTALHHSVY--RECSAGHVLQVLTCT--ACGHR--RGSSRIVGNNMSLQSPWQAS	233
Db	818	LYFDQQVQCGASLVSRDWLVSAACVYGRNMEPSKWKAVLGLHMASNLSPQIERLIDQ	877	QY	290	IVYHRYKPKLGNLALKMKGAP--PKNEMQPOVLPNSBNFPGKVCWMSKGWQADEG	349	QY	290	IVYHRYKPKLGNLALKMKGAP--PKNEMQPOVLPNSBNFPGKVCWMSKGWQADEG	349
Db	878	IVINPHYNKRKNNDLAMMHEMLKMYNTDYLQICLPEENOFPGRLCISAGLMAH	936	QY	350	AGDASVFLNHAAPVPLJSNKCNHR-DVYGGIISPSLCAQYLTGyvDSCGDSGPVLCQ	408	QY	350	AGDASVFLNHAAPVPLJSNKCNHR-DVYGGIISPSLCAQYLTGyvDSCGDSGPVLCQ	408
Db	937	OGSTDADVLQEDAVPLLSNEKQCOQHPEYN--ITENNCAGYEAGGVDSQCGDGGPLMCQ	994	Db	937	937 OGSTDADVLQEDAVPLLSNEKQCOQHPEYN--ITENNCAGYEAGGVDSQCGDGGPLMCQ	994	Db	937	937 OGSTDADVLQEDAVPLLSNEKQCOQHPEYN--ITENNCAGYEAGGVDSQCGDGGPLMCQ	994
Db	995	ENNRWILLAGVTSFGQCALPNRPGVYARVPRFTENI	1030	Db	995	995 ENNRWILAGVTSFGQCALPNRPGVYARVPRFTENI	1030	Db	995	995 ENNRWILAGVTSFGQCALPNRPGVYARVPRFTENI	1030
RESULT	3										
A56318		enteropeptidase (EC 3.4.21.9) precursor - human									
N;Alternate names: enterokinase											
C;Species: Homo sapiens (man)											
C;Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 18-Jun-1999											
C;Accession: A56318; B4390											
R;Kitamoto, Y.; Veille, R.A.; Donis-Keller, H.; Sadler, J.E.											
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994											
A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of intestinal enteropeptidase and trypsin.											
A;Reference number: A56318; MUID:95234679											
A;Accession: A56318											
A;Molecule type: mRNA											
A;Residues: 1-1019 <K1T>											
A;Cross-references: GB:U09860; PIDN:9746412; PIDN: AAC50138.1; PID:9746413											
R;Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.											
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994											
A;Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of intestinal enteropeptidase and trypsin.											
A;Reference number: A43090; MUID:94329561											
A;Accession: B43090											
A;Status: nucleic acid sequence not shown											
A;Molecule type: mRNA											
A;Residues: 745-1019 <K1T>											
A;Cross-references: GB:U09860											
C;Comment: The mechanism of association with the membrane of the intestinal brush border membrane attachment using a signal-anchor sequence.											
C;Comment: Conversion from membrane-bound to soluble forms may involve further processing.											
C;Complex: mature enteropeptidase is variously reported to contain two (heavy and light) isoflde linked											
C;Function:											
A;Gene: GDB: PRS7											
A;Cross-references: GDB:384083; OMIM:226200											
A;Map Position: 21q21-21q21											
C;Complex: Mature enteropeptidase is variously reported to contain two (heavy and light) isoflde linked											
C;Comment: The mechanism of association with the membrane of the intestinal brush border membrane attachment using a signal-anchor sequence.											
C;Comment: Conversion from membrane-bound to soluble forms may involve further processing.											
C;Complex: mature enteropeptidase is variously reported to contain two (heavy and light) isoflde linked											
C;Function:											
A;Description: cleaves activation peptide from trypsinogen to produce active trypsin											
A;Pathway: intestinal digestive hydrolytic cascade											
A;Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding repeat homology											
C;Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein											
F;52-117/product: enteropeptidase mini chain #status predicted <MCH>											
F;118-800/product: enteropeptidase heavy chain #status predicted <RCH>											
F;199-236/domain: LDL receptor ligand-binding repeat homology <LDL1>											
F;358-520/domain: MAM homology <MAM>											
F;542-647/domain: C1r/C1s repeat homology <C1R>											
F;659-683/domain: LDL receptor ligand-binding repeat homology <LDL2>											
F;694-799/domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC>											
F;801-1035/product: enteropeptidase light chain #status predicted <CLH>											
F;116-147,170,194,233,263,264,404,456,519,550,646,698,722,741,762,864,903,965/Bindin											
F;788-912,826-842,926-993,957-972,981-1011/Disulfide bonds: #status predicted											
F;841,892,987/Active site: His, Asp, Ser #status predicted											
Query Match 27.0%; Score 660; DB 1; Length 1035;											
Best Local Similarity 36.9%; Pred. No. 1e-44;											
Matches 146; conservative 70; Mismatches 146; Indels 34; Gaps 15;											
Db	603	603-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>	120	QY	66	GLGIFHDC-SGKYRCRSSFCIILARCDGVSDCKDGEDEYCVRV---GGONAVLQWF	120	QY	66	GLGIFHDC-SGKYRCRSSFCIILARCDGVSDCKDGEDEYCVRV---GGONAVLQWF	120
Db	678	678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC>	120	QY	652	GLGIPPECKEDNFOCKDG-ECIPLVNLCDGDFPHCKDGSDBAHCVRLFNFTDSSGLVQFF	710	QY	652	GLGIPPECKEDNFOCKDG-ECIPLVNLCDGDFPHCKDGSDBAHCVRLFNFTDSSGLVQFF	710

F;72-896,810-826,910-977,941-956,967-995/disulfide bonds: #status predicted
F;825,876,971/Active site: His, Asp, Ser #status predicted

F;787-911,825-841,925-992,956-971,982-1010/disulfide bonds: #status predicted
F;840,891,986/Active site: His, Asp, Ser #status predicted

Query Match 26.8%; Score 657; DB 1; Length 1019;
Best Local Similarity 37.7%; Pred. No. 1..7e-44;
Matches 148; Conservative 67; Mismatches 148; Indels 30; Gaps 15;
Db 637 LGIPECKADHFQCKNG-ECVPLVNLCDGHHLACEDSDEADCVRFNGTTNNNGLVRFRI 695

Qy 67 LGIHFDC-SGKYRCRSSFKCIELIARCDGVSDCKDGEDEYRCVR---VGGQNAVQVFT 121
Db 652 LGIPECKEDNFQCEG-ECVLLVNLCDGHSHCKDGSDEAHCVRLFLNGTANNGLVWFR 710

Qy 122 AASWKMCSDDMKGHVANVACQALGPSPVSSDNLRVSSLEGQFRREFVSDHLLPDKV 181
Db 696 QSIWHTACAEWNTTQTSNDVCPOLLGLGSGNSSK- PIFSTDG--GPFKVL-NTADG -- 747

Qy 182 TALHHSYVVRGCGASGHVWVLOCT--ACGHR--RGYSSRIVGGNMSSLSSOWPWAQSLQF 236
Db 748 --HHLTPSQCQLQDLSLRLQCNHRSCKGKLLAQQDITPKIVGSMNAKEGAWPWWVGLY 804

Qy 237 QGYHLCGGSVITPLWITTAACVYDYL-PKSWTQVGL--VSLIDNP-APSHLVEKIV 292
Db 805 GGRLLCGASLVSQDWLVAACVYGRNLEPSKWTAILGLHMKSLNTSPQIVPRLDEI 864

Qy 293 HSXKVKRIGNDIAIMKLAGLTLFENEMIOPVCLPNSENFDPGKVWTSGWATGAGD 352
Db 865 NPHYNRKRDNDIAMMHLLEFKVNTDQPICLPENQVTPGRNCSTAGWG-TVWYQGT 923

Qy 353 ASPVLNVNHAVLNSKICNHR-DVYGGITPSMLCAGYLPGVDSQGDGGPLMCQENN 981
Db • 924 TANLQEADVPPLSNRCQDQMPFYN- ITENMCMAGYEGGIDSCQGDGGPLMCQENN 981

Qy 412 LWKLGVATSGFGIGCAEVNKPVYVTRTSFLWI 444
Db 982 RWFLAGVTSFGYKCALPNRPGVYARVSRFTEWI 1014

RESULT 4

A53663 enteropeptidase (EC 3.4.21.9) precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Accession: A53663
R;Matushima, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada, S.; Miki, K.; Kurokawa, J.; Biol. Chem. 269 19976-19982, 1994
A;Title: Structural characterization of porcine enteropeptidase.
A;Reference number: A53663; MUID: 9427548
A;Accession: A53663
A;Molecule type: mRNA
A;Residues: 1..1034 <RAT>
A;Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123
A;Note: Parts of this sequence, including the amino ends of three chains isolated from t
C;Comment: The mechanism of association with the membrane of the intestinal brush border (stated below) or with amino-terminal myristylation of the heavy chain.
C;Complex: Mature enteropeptidase is variously reported to contain two (heavy and light) ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
C;Function:
A;Description: cleaves activation peptide from trypsinogen to produce active trypsin
A;Pathway: intestinal digestive hydrolase cascade
C;Superfamily: enteropeptidase; C1r/C1s repeat homology; LDL receptor ligand-binding rep
C;Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F;22-38/Domain: transmembrane #status predicted <MM>
F;5-21/7/Product: enteropeptidase mini chain #status predicted <MM>
F;18-799/Product: enteropeptidase heavy chain #status predicted <MM>
F;19-236/Domain: LDL receptor ligand-binding repeat homology <MM>
F;357-519/Domain: MAM homology <MM>
F;541-646/Domain: C1r/C1s repeat homology <C1R>
F;658-692/Domain: LDL receptor ligand-binding repeat homology <DLR>
F;693-798/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC
F;800-1034/Product: enteropeptidase light chain #status predicted <MM>
F;800-1029/Domain: trypsin homology <TRI>
F;116,147,170,194,283,343,350,403,455,485,518,549,645,697,701,721,740,761,804,863,902,98

RESULT 5

S3377 hepsin (EC 3.4.21. -) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Accession: S3377; S32013
R;Farley, D.; Raymond, F.; Nick, H.
Biochim. Biophys. Acta 1173, 350-352, 1993
A;Title: Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase
A;Reference number: S3377; MUID: 93305733
A;Accession: S3377
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1..416 <RAT>
A;Cross-references: EMBL:X70900; NID:957928; PIDN:CAA50256.1; PID:957929
C;Superfamily: hepsin; trypsin homology
C;Keywords: hydrolase; liver; serine proteinase; transmembrane protein
F;22-24/Domain: transmembrane #status predicted <MM>
F;162-399/Domain: trypsin homology <TRI>
F;187-203,290-358,312-337,348-380/disulfide bonds: #status predicted
F;202,256-352/Active site: His, Asp, Ser #status predicted

Query Match 26.4%; Score 645; DB 1; Length 416;
Best Local Similarity 35.2%; Pred. No. 5..5e-44;
Matches 146; Conservative 61; Mismatches 52; Indels 52; Gaps 11;
Db 60 ILALAGLIGHFDCSKYCRSSFKCIELIARCDGVSDCKDGEDEYRCVYGGQNAVQV 119
Qy 177 VAALTVGTLFLPTGIG---AASWAVTILR-----SDQEPYQVQSPGSQRLV 65

Qy 120 -FTAAWSKTMCSDDWKGHVANVACQQLGPFSYVSSDNLRVSSLEGQFREEFVSD---- 173
Db 66 DKTEGIVRLQSSRSNARVAGLGCSEMGFJRALHSELDVRTAGANGTSQFFCDEGGP 125
Qy 174 ---HLLPDDKVITALHHSVYVREGCASGHVWVLOCTACGHRGGYSSRIVGGNMSLLSQWPW 230

Db	301 DFEGEELNV-----TFVQGADVCOETCTKTRICOFF-IVSLLPOD-----CKE 342	F;127-308,396,453,494/Binding site: carbohydrate (asn) (covalent) #status experimenta
Qy	192 EGC-----ASCHVYL-----QCTAGHRRGYSRRIVGGMMS 223	F;390-391/Cleavage site: Arg-ile (coagulation factor XIA) #status predicted
Db	343 EGCCKCSRLLSTDGSPTRITYGMGSSYSSRLCKYDSDPCIT-----KINARIVGNTA 397	F;434,483,518/Active site: His, Asp, Ser #status predicted
Qy	224 LIISQWPMQASLQFQ---GYHLCGGSVITPLWILITAACVYDILPKSWTIQVGLVSI-L 278	
Db	398 SLGEMWPMQVSQLVKLYQSQTQHLOGSLSIGRQWLTAAHCFDGFIPDWPWRIKGILSLEI 457	
Qy	279 DNAPSHLVEKIVYHVKYKPKRKLNDALMLKLAGPLTFNEMIQPVCLPNSNENPDKV 338	
Db	458 TKETPSSRRIKELIHQEYKVSEGNYDIALIKLQLQPLVTEFORKPCLPSKAINTNTYTC 517	
Qy	339 WTSGGATEDGAGDASVNLNHAAPLISNKC---NHRDVGGIISMSMLCAGYLTCGVDS 396	
Db	518 WYWGWTKE-OGETONILKATIPLVNPNECQKRYD---VINKOMICAGYKEGTD 573	
Qy	397 CGDDSGPLVCOBERRLWLVGATTSFGTGCAEVNKGPGVYTRVTSFLDWIHEQME---RDL 452	
Db	574 CGDDSGPLVCOBERRLWLVGATTSFGTGCAEVNKGPGVYTRVTSFLDWIHEQME---RDL 633	
Qy	453 KT 454	
Db	634 ET 635	
RESULT 8		
QKHP		
Plasma kallikrein (EC 3.4.21.34) precursor - human		
C;Alternate names: Kininogen; Plasma prekallikrein		
C;Species: Homo sapiens (man)		
C;Accession: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 18-Jun-1999		
R;Chung, D.W.; Fujikawa, K.; McMullen, B.A.; Davie, E.W.		
Biochemistry 25, 2410-2417, 1986		
A;Title: Human plasma prekallikrein, a zymogen to a serine protease that contains four t		
A;Reference number: A00921; MUID:86243359		
A;Molecule type: mRNA		
A;Residues: 1-638 <CHD>		
A;Cross-references: GB: M13143; NID:9190262; PTDN:AAA60153.1; PID:9190263		
R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.		
Biochemistry 30, 2050-2056, 1991		
A;Title: Location of the disulfide bonds in human plasma prekallikrein: the presence of		
A;Reference number: A37939; MUID:91152016		
A;Accession: A37939		
A;Molecule type: protein		
A;Residues: 20-27,40-46, 'X', '48', 'H', '50', 'X', '52-70, 'H', '75-76, 'X', '78-80; 103-113; 131-140,141-260-203, 'X', '285,287-291, 'X', '293-295,314,317, 'X', '319-320,321-324; 'X', '329-333; 334-339, 'X', '525,538-551,562, 'X', '564-567,573, 'X', '575-576,578-583, 'X', '585,592-604 <NCB>		
C;Comment: This protein, synthesised in the liver, circulates as a noncovalent complex w		
C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li		
C;Comment: The enzyme cleaves Lys-Arg and Arg-Ser bonds. It activates, in a reciprocal r		
ingen and may also play a role in the renin-angiotensin system by converting prorenin i		
C;Genetics:		
A;Gene: GDB:KLK3		
A;Cross-references: GDB:127575; OMIM:229000		
A;Map position: 4q35-4q35		
C;Superfamily: coagulation factor XI; trypsin homology		
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla		
F;1-19/Domain: signal sequence #status predicted <SIG>		
F;20-538/Domain: plasma kallikrein #status predicted <NAT>		
F;20-90/Domain: plasma kallikrein heavy chain #status predicted <HCH>		
F;20-109/Domain: apple repeat <AP1>		
F;110-199/Domain: apple repeat <AP2>		
F;200-289/Domain: apple repeat <AP3>		
F;291-380/Domain: apple repeat <AP4>		
F;391-638/Domain: plasma kallikrein light chain #status predicted <LCH>		
F;391-621/Domain: trypsin homology <TRY>		
F;21-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,322-328,383		
RESULT 9		
PLPG		
Plasmin (EC 3.4.21.7) precursor - pig (fragment)		
N;Alternate names: plasminogen		
N;Contains: miniplasminogen		
C;Species: Sus scrofa domestica (domestic pig)		
C;Date: 07-Sep-1990 #sequence_revision 01-Nov-1996 #text_change 18-Jul-1997		
C;Accession: S07733; S03737; A52834		
R;Schaller, J.; Marti, T.; Roesselet, S.J.; Kaempfer, U.; Rickli, B.E.		
Eur. J. Biochem. 114, 465-470, 1981		
R;Schaller, J.; Marti, T.; Roesselet, S.J.; Kaempfer, U.; Rickli, B.E.		
A;Title: Comparison of the primary structure of the N-terminal CNBr fragments of huma		
A;Reference number: S03733		
A;Accession: S03733		
A;Molecule type: protein		
A;Residues: 1-560 <SCH>		
R;Brunisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg,		
Eur. J. Biochem. 114, 465-470, 1981		
A;Title: Comparison of the primary structure of the N-terminal CNBr fragments of huma		
A;Reference number: S03735; MUID:81212097		
A;Accession: S03737		
A;Molecule type: Protein		
A;Residues: 1-57 <BRU>		
R;Marti, T.; Schaller, J.; Rickli, E.E.		
Eur. J. Biochem. 149, 279-285, 1985		
A;Title: Determination of the complete amino-acid sequence of porcine miniplasminogen		
A;Reference number: A25834; MUID:85203907		
A;Accession: A25834		
A;Molecule type: protein		
A;Residues: 450-790 <MAR>		
C;Function: dissolves the fibrin of blood clots; acts as a proteolytic factor in a		
A;Pathway: fibrinolysis		
C;Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homol		
C;Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; plasma; serine pr		
F;1-77/Domain: plasminogen-related protein precursor homology (fragment) <PLPH>		
F;1-77/Domain: activation peptide #status predicted <APT>		

F:78-560/Product: plasmin chain A #status predicted <ACH>
 F:84-162/Domain: kringle homology <KR1>
 F:114-243/Domain: kringle homology <KR2>
 F:256-333/Domain: kringle homology <KR3>
 F:358-435/Domain: kringle homology <KR4>
 F:450-590/Product: miniplasminogen #status experimental <MIN>
 F:461-540/Domain: kringle homology <KR5>
 F:561-790/Product: plasmin chain B #status experimental <BCH>
 F:30-54-34-42, 84-162, 105-145, 133-157, 166-243, 169-297, 187-226, 215-238, 256-333, 277-316, 305 bonds: #status predicted
 F:602, 645, 740/Active site: His, Asp, Ser #status predicted

Query Match 21.5%; Score 526; DB 1; Length 790;

Best Local Similarity 41.7%; Pred. No. 3, 8e-34; Matches 110; Conservative 42; Mismatches 86; Indels 26; Gaps 7;

QY 194 CASGHVWVLOCTTCGHRGRGSSIVGGNMSLISQWPWQASLQFQGY --HLCGGSVITPLW 251

Db 547 CGPKVEPKKCPA-----RNVGGCVSIPHSWQWISLRY-YRGRHFCGGTLISPEW 596

QY 252 IITRAHCYDLYPKSWMTIQVGLVSLNDNPAPSHLVERI -VYHSKKPKRKLGGNDIALMK 309

Db 597 VTAHKCLEKSSSPSSVYK-----IIGAHEEYHHLGEQVQETDVKLFKEPSEADIALLK 650

QY 310 LAGPLTFNEMIQVCLPNEENFPDPKGKVWTSGWATED -GAGDASPVLNHAAPLTSN 367

Db 651 LSSPAVITDKVIRACLPFPNVVADRTACYIQTGGTGTGAG---LLKEARLPVIE 706

QY 368 KTCNHRDVTGGGIPSMSMCAGY-TGGVDSQCGNSGGGPPVQCOERLWKVUGVATSGIGCAE 427

Db 707 KVCNRYEYIGGKVSPELNLCAHLAGGIDSCQGSDGGPLVCFEHDKYILQGVTSWGLGAL 766

QY 428 VNPQGVWYRVTSLWQHMQED 451

Db 767 PNRPGVYVVRSPRVTWIEIMRRN 790

RESULT 10
 JC5759

brain-specific serine proteinase (EC 3.4.21.-) - mouse

C:Species: mus musculus (house mouse)
 C:Accession: JC5759

R.Yamamoto, Y.; Yamashiro, K.; Tsuruoka, N.; Nakazato, H.; Tsujimura, A.; Yamaguchi, N.

Biochem. Biophys. Res. Commun., 239, 386-392, 1997

A;Title: Molecular cloning of a novel brain-specific serine protease with a kringle-like

A;Reference number: JC5759; MUID:98008848

A;Molecule type: mRNA

A;Cross-references: DDBJ:d89871

A;Experimental source: brain

C:Superfamily: brain-specific serine proteinase; scavenger receptor cysteine-rich domain

C:Keywords: glycoprotein; hydrolase; serine proteinase

F:85-177/Domain: kingle-like #status predicted <KR1>

F:116-266/Domain: scavenger receptor cysteine-rich domain homology <SRC7>

F:216-266, 273-372, 386-486/Domain: scavenger receptor cysteine-rich #status predicted <SR>

F:513-516/Domain: furin binding #status predicted <PRB>

F:517-515/Domain: trypsin homology <TRY>

F:521-569/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:562, 612, 711/Active site: His, Asp, Ser #status predicted

Query Match 21.4%; Score 523.5; DB 2; Length 761;

Best Local Similarity 32.0%; Pred. No. 5, 8e-34; Matches 148; Conservative 53; Mismatches 141; Indels 121; Gaps 22;

QY 70 HFD-----CSGKYRCRSSFCIELTARDGVSDFCKGED----- 103

Db 326 HDGSNNRPIWLDVSCSGK--EVSF--IQCRRQWGRHDCSHREDVGLCYPFDSDCHRL 380

Query Match 21.3%; Score 520.5; DB 2; Length 455;

Best Local Similarity 41.4%; Pred. No. 5, 5e-34; Matches 106; Conservative 48; Mismatches 87; Indels 15; Gaps 6;

QY 203 OCTA---CG---HRRQYSSRVGGNMSLISQWPWQASLQFQ-GHLLGGSVITPLW 253

Db 204 QCSSPFDGPKVPEPKRGSRVGGVLAHSPWQWISLRPGRFCGGPLISPEW 263

QY 254 TAAHCYDLYPKSWMTIQVGLVSLNDNPAPSHLVEKIVYHSKKPKRKLGGNDIALMLAGP 313

Db 264 TRAACLERSRSPSTKVLGLTHHELRLAAGQQID--VSKLFLEPSRA--DIALLKLSSP 319

Qy 314 LTFNEIQPVCPLPNEENFDGKWCNTSGNATEDGAGDASPVLNHAAPLISKICNHR 373

Db 320 AIIITONVIAPCLPAPDVVANWAECFVTGMGETQDSNN--AGVIVAEQVIEENKVCNRY 377

Qy 374 DVGIGITSPSMICAGYLGGVDSQCGDGGPLVCOERRLWKLVAGATSGFGCAEVNKGPGV 433

Db 378 BYLNGRKVSTELCAGHLVGGVSDCOGDSGGPLVFEKDKYILOQVTSWGLGCARPNSPGV 437

Qy 434 YTRVYNSFLDWMHEME 449

Db 438 YVRVSSFINNTERIMQ 453

RESULT 12

plasmin (EC 3.4.21.7) precursor - mouse

N;Contains: angiotatin; plasminogen

C;Species: Mus musculus (house mouse)

C;Date: 20-Sep-1991 #sequence_revision 01-Nov-1996 #text_change 18-Jun-1999

C;Accession: A38514; S48202; S48203

R;Deen, S.J.F.; Bell, S.M.; Schafer, L.A.; Elliott, R.W.

Genomics 8: 49-61, 1990

A;Title: Characterization of the cDNA coding for mouse plasminogen and localization of t

A;Accession: A38514

A;Molecule type: mRNA

A;Residues: 1-612 <DB>

A;Cross-references: GB:J04766; NID:9200402; PID:AAA50168.1; PID:9200403

R;Lijnen, H.R.; van Hoei, B.; Beelen, V.; Collen, D.

Eur. J. Biochem. 224: 863-871, 1994

A;Title: Characterization of the murine plasma fibrinolytic system.

A;Reference number: S48202; MUID:95010076

A;Accession: S48202

A;Molecule type: protein

A;Residues: 20-25 <DB>

A;Accession: S48203

A;Molecule type: protein

A;Residues: 22-27 <DB>

C;Comment: Plasminogen is synthesized by the kidney and is present in plasma and many of

medically after dissociation from the clot. In the presence of the inhibitor, the activa

e inhibitor, the activation involves also removal of the activation peptide.

C;Comment: Stromelysin 1 (see PIR:KCMS1) acts on plasminogen to produce angiotatin. Tc

C;Function:

A;Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va

ns the walls of the graafian follicle; also activates the urokinase type plasminogen act

A;Pathway: fibrinolysis

C;Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology <PLPH>

F;1-96/Domain: plasminogen-related protein precursor homology <PLPH>

F;1-19/Domain: signal sequence #status predicted <SIG>

F;2-81/2/Product: plasminogen #status predicted <PRO>

F;2-96/Domain: activation peptide #status predicted <APT>

F;7-9-466/Product: angiotatin #status predicted <AST>

F;9-581-582-812/Product: plasmin #status predicted <PAT>

F;97-581/Domain: chain A #status Predicted <ACH>

F;10-181/2/Domain: kringle homology <KR1>

F;18-262/2/Domain: kringle homology <KR2>

F;25-352/2/Domain: kringle homology <KR3>

F;377-454/2/Domain: kringle homology <KR4>

F;481-560/2/Domain: kringle homology <KR5>

F;582-812/2/Domain: chain B #status predicted <BCB>

F;49-73, 55-61/103-181, 124-164, 152-176, 185-252, 188-316, 206-245, 234-257, 275-352, 296-335, 32

bonds. #status predicted

F;78-79/Cleavage site: Gu-Asn (stromelysin 1) #status predicted

F;136/238/Bindng site: carbonyl (Asn) (covalent) #status predicted

F;465-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted

F;581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental

F;624, 667, 762/Active site: His, Asp, Ser #status predicted

Query Match 20.7%; Score 507.5; DB 1; Length 812;

Best Local Similarity 39.8%; Pred. No. 1.2e-32; Matches 106; Conservative 40; Mismatches 91; Index 29; Gaps 7;

Db 568 CGKPOVEPKC-----PGRVGGCVANPHSPWQPOISLRTRFTGQHFCGGTLIAPW 618

Qy 252 LITAHCYVQDYLPLPSWTLQG---LVSLLDNPARSHVIEKIVHSKPKRIGNDL 307

Db 619 VLTAAHCLEKSSRPEFYKVILGAHEEYVIRGLDVOITS-VAKLILEPN----NRDIAL 670

Qy 308 MKLAGPLTFNEMIQPVCLPMPSEENFPDGKWCNTSGNATEDGAGDASPVLNHAAPL 365

Db 671 LKLSRIPATPTDKVPLCPLSPPNMYADRTCPYIICLWGEPTGFGR---LKERRQPLV 725

Qy 366 SNKICNRHDYGGITSPSMICAGYLGGVNSCQGDGGPPTVCQERRLWKLVAGATSGFGPGC 425

Db 727 ENKVCNRVEFLNIRVKSTELCAGQLLAGVNSCQGDGGPLVCFERDKYILOQVTSWGLGC 786

Qy 426 AEVNPVGVYRVTSTLDWIBQMER 454

Db 787 ARPNPKPGVYVRSRFWDWIREMRNN 812

RESULT 13

acrosin (EC 3.4.21.10) precursor - rat

N;Contains: proacrosin

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999

C;Accession: S18407; S30037; M56620

R;Kleinn, U.; Flake, A.; Engel, W.

Biochim. Biophys. Acta 1090: 270-272, 1991

A;Title: Rat Sperm acrosin: cDNA sequence, derived primary structure and phylogenetic

A;Accession: S18407

A;Molecule type: mRNA

A;Residues: 1-437 <RLE>

A;Cross-references: EMBL: X59354

R;Kleinn, U.; Flake, A.; Engel, W.

A;Title: Submission to the EMBL Data Library, April 1991

A;Reference number: S30037

A;Accession: S30037

A;Molecule type: mRNA

A;Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KLE2>

A;Cross-references: EMBL:X59354; NID:957282; PID:CAAN1947.1; PID:957283

R;Kramling, H.; Flake, A.; Adham, I.M.; Radtke, J.; Engel, W.

DNA Seq. 2: 57-60, 1991.

A;Title: Exon-intron structure and nucleotide sequence of the rat proacrosin gene.

A;Reference number: A56620; MUID:92199245

A;Accession: A56620

A;Status: preliminary

A;Molecule type: DNA; mRNA

A;Residues: 1-254, 'ICDR', 259, 'DHEL', 264, 'GRIC', 269-437 <KRE>

A;Note: Sequence modified after extraction from NCBI backbone (NCBIN:89436, NCBIN:89439, NCBIN:89447, NCBIN:89447, NCBIN:89447)

C;Superfamily: acrosin; trypsin homology

C;Keywords: glycoprotein; hydrolase; serine proteinase; sperm; zymogen

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-812/Domain: activation peptide #status predicted <PRO>

F;2-96/Domain: activation peptide #status predicted <APT>

F;7-9-466/Domain: angiotatin #status predicted <AST>

F;9-581-582-812/Domain: plasmin #status predicted <PAT>

F;97-581/Domain: chain A #status Predicted <ACH>

F;10-181/2/Domain: kringle homology <KR1>

F;18-262/2/Domain: kringle homology <KR2>

F;25-352/2/Domain: kringle homology <KR3>

F;377-454/2/Domain: kringle homology <KR4>

F;481-560/2/Domain: kringle homology <KR5>

F;582-812/2/Domain: chain B #status predicted <BCB>

F;49-73, 55-61/103-181, 124-164, 152-176, 185-252, 188-316, 206-245, 234-257, 275-352, 296-335, 32

bonds. #status predicted

F;78-79/Cleavage site: Gu-Asn (stromelysin 1) #status predicted

F;136/238/Bindng site: carbonyl (Asn) (covalent) #status predicted

F;465-467/Cleavage site: Thr-Val (stromelysin 1) #status predicted

F;581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental

F;25-159-216/163/disulfide bonds: #status predicted

F;74-90/disulfide bonds: #status predicted

F;89-143-241/Active site: His, Asp, Ser #status predicted

F;178-247/Disulfide bonds: #status predicted

F;210-226/Disulfide bonds: #status predicted

F;237-267/Disulfide bonds: #status predicted

Query Match	20.7%	Score 507;	DB 2;	Length 437;
Best Local Similarity	42.4%	Pred. No. 5..3e-33;		
Matches 104;	Conservative	Mismatches 89;	Indels 16;	Gap 5;
Qy	216 RIVGGNMNLLSQWPWQASLQF----	Score 507;	DB 2;	Length 437;
Db	42 RIVGGTQHSSRWRWPWMYSLQITSHNSRRYHACGGSILNSHNLTAACFDNKVVQDW	Score 507;	DB 2;	Length 437;
Qy	264 LP-KSWTIQVGLLSDNPAPSHLVEKIVHKKYKPKRLGNDIALMPLLAGPLTFNEMIQ	Score 507;	DB 2;	Length 437;
Db	102 LVFGAHELEYGRNKPVREPOQFRRVQKIVIRKKYNAVEGNDIALLKVTPPTCGPGVGP	Score 507;	DB 2;	Length 437;
Qy	323 VCLPNSEENFPD-GKVWTSGMATEDGAGDASPVLNHAAPLISNKICNRDVGCGIS 381	Score 507;	DB 2;	Length 437;
Db	162 GCLPHFKSGPPRPHCTVIGVYKDNAPRSPSPVLMEARVDLIDLCNSIQWYNGRV 221	Score 507;	DB 2;	Length 437;
Qy	382 PSMICAGYLTSQGDSGQPLVQCE--RLWKLVLGATSGIGAEVNPGVYRTVS 439	Score 507;	DB 2;	Length 437;
Db	222 STHVCAGSPEKGKTDQGDGGPLMCRDTRROFVIGVTSWGVGCARAKRPGVYATWD 281	Score 507;	DB 2;	Length 437;
Qy	440 FIDWI 444	Score 507;	DB 2;	Length 437;
Db	282 YIDWI 286	Score 507;	DB 2;	Length 437;
RESULT 14				
JE0315	low-density lipoprotein receptor-related protein - mouse			
C:Species	Mus musculus (house mouse)			
C:Date	16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 18-Aug-2000			
C:Accession	JE0315			
R:Tomita, Y.; Kim, D.; Maggori, K.; Fujino, T.; Yamamoto, T.T.				
J:Biochem. 124, 784-789, 1998				
A:Title: A novel low-density lipoprotein receptor-related protein with type II membrane				
A:Reference number: JE0315; MUID:98429596				
A:Accession: JE0315				
A:Status: preliminary				
A:Residues: 1-113 <TOM>				
A:Cross-references: DDBJ:AB013874; NID:93869144; PIDN:BA34371.1; PID:93869145				
A:Molecule type: mRNA				
A:Residues: 1-113 <TOM>				
A:Cross-references: DDBJ:AB013874; NID:93869144; PIDN:BA34371.1; PID:93869145				
C:Superfamily: trypsin homology; LDL receptor ligand-binding repeat homology <LDL1> F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2> F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3> F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4> F:648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5> F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6> F:721-750/Domain: LDL receptor ligand-binding repeat homology <LDL7> F:869-1097/Domain: trypsin homology <TRY>				
Query Match	20.7%	Score 506;	DB 2;	Length 113;
Best Local Similarity	31.7%	Pred. No. 2.2e-32;		
Matches 126;	Conservative	Mismatches 66;	Indels 36;	Gap 16;
Qy	72 DCS---GKYRGRSSFKCIELIARCDGVSDKDEDEYRCRV--GGONAVLQVTAAS 124	Score 506;	DB 2;	Length 113;
Db	719 NCSCQDNELEC-ANHCVPRQLWCDGWDVCDSSDEWGCYILSKKNSSSLLVHSAK 777	Score 506;	DB 2;	Length 113;
Qy	125 WKTMCSDWKHYANVACAOLGF--PSYVSSDNLYRSSLEGOFREEVSDIHLPPDKV 182	Score 506;	DB 2;	Length 113;
Db	778 EHHVCADGWRETIQLACKOMGIGEPSEV-----TKLIPQEGQOWLRLYPNWENLNGS 830	Score 506;	DB 2;	Length 113;
Qy	183 ALHHSVYVREGGASA GHVWLQCTA-CGHRG--YSSRIVGGMMSLISOWWQASLQFO- 237	Score 506;	DB 2;	Length 113;
Db	831 TQELLYVYRHSQPSRSBISLCKQDGRRPAARMKRILGGRTSRPGWVWQCSQSE 890	Score 506;	DB 2;	Length 113;
Qy	238 GHLCGGSVITPLWITAAHCYVVDLYLPKSWTIVQVLSLDDNA---PSHLVKEVYHS 294	Score 506;	DB 2;	Length 113;
Db	891 SGHICGVLIAKKWLTVAHCEGREGEDADWVWVFG-INNLDPHPSGEMOTRFRVKTILHP 949	Score 506;	DB 2;	Length 113;
RESULT 15				
PLRU				
plasmin (EC 3.4.21.7) precursor [validated] - human				
M:Alternative names: plasminogen precursor [mismomer]				
N:Contains: angiotensin; plasminogen				
C:Species: Homo sapiens (man)				
C:Date: 24-Apr-1984 #sequence_revision 02-Dec-1994 #text_change 15-Sep-2000				
C:Accession: A35229; 152242; A2646; 162138; I84609; S03735; A00929; A04627; A04625;				
R:Petersen, T.E.; Marzen, M.R.; Ichinose, A.; Davie, E.W.				
J: Biol. Chem. 265, 6104-6111, 1990				
A:Title: Characterization of the gene for human plasminogen, a key proenzyme in the f				
A:Reference number: A35229; MUID:90202879				
A:Accession: A35229				
A:Molecule type: DNA				
A:Residues: 1-810 <PET>				
A:Cross-references: GB:J05286; GB:MB4276; GB:MB4277; PIDN:AAA60113.1; PID:9387026				
A:Experimental source: leukocyte; lung fibroblast				
R:Malagrétti, N.; Bruno, L.; Pontoglio, M.; Condiani, G.; Meroni, G.; Ottolenghi, S.; Biochem. Biophys. Res. Commun. 173, 1013-1018, 1990				
A:Title: Definition of the transcription initiation site of human plasminogen gene in				
A:Reference number: I52242; MUID:91097523				
A:Status: translated from GB/EMBL/DDBJ				
A:Molecule type: DNA				
A:Residues: 1-16 <MAL1>				
A:Cross-references: GB:MB2690; NID:9190097; PIDN:AAA36454.1; PID:9553613				
R:Forssén, M.; Baden, B.; Israelsson, M.; Larsson, K.; Heden, L.O.				
FBBS Lett. 213, 254-260, 1987				
A:Reference number: A26646; MUID:87162490				
A:Accession: A26646				
A:Molecule type: mRNA				
A:Residues: 1-471; D, 473-810 <FOR>				
A:Cross-references: GB:X05199; NID:935530; PIDN:CAA28831.1; PID:935531				
A:Experimental source: liver				
R:Malinowski, D.P.; Sadler, J.E.; Davie, E.W.				
Biochemistry 23, 421-425, 1984				
A:Title: Characterization of a complementary deoxyribonucleic acid coding for human a				
A:Reference number: 145961; MUID:85023311				
A:Status: translated from GB/EMBL/DDBJ				
A:Accession: 162738				
A:Molecule type: mRNA				
A:Residues: 292-471; D, 473-810 <MAL2>				
A:Cross-references: GB:K02922; NID:9190112; PIDN:AAA60124.1; PID:9387031				
A:Accession: 184609				
A:Status: translated from GB/EMBL/DDBJ				
A:Molecule type: DNA				
A:Residues: 367-419 <MAL3>				
A:Cross-references: GB:K02921; NID:9190110; PIDN:AAA60123.1; PID:9190111				
R:Brünisholz, R.A.; Lierch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, Eur. J. Biochem. 114, 465-470, 1981				
A:Title: Comparison of the primary structure of the N-terminal CNBr fragments of huma				
A:Reference number: S03735; MUID:81212097				
A:Accession: S03735				
A:Molecule type: protein				
A:Residues: 20-71; E, 73-76 <BR0>				
R:Søttrup-Jensen, L.; Petersen, T.E.; Magnusson, S.				
A:Title: submitted to the Atlas, July 1977				
A:Reference number: A00929				
A:Accession: A00929				

A; Molecule type: protein
 A; Residues: 20-71, 'E', 73-85, 87-106, 'D', 108-360, 'E', 362-810 <50T>
 R; Wiman, B.; Biochem. 76, 129-137, 1977
 A; Title: Primary structure of the B-chain of human plasmin.
 A; Reference number: A04627; MUID:7725245
 A; Accession: A04627;
 A; Molecule type: protein
 A; Residues: 581-810 <W11>
 R; Wiman, B.; Wallen, P.; Biochem. 50, 489-494, 1975
 A; Title: Structural relationship between "glutamic acid" and "lysine" forms of human plasminogen that
 A; Reference number: A04625; MUID:75093329
 A; Accession: A04625
 A; Molecule type: protein
 A; Residues: 20-50, 'Q', 51-71, 'E', 73-85, 87-100 <W12>
 R; Wiman, B.; Wallen, P.; Biochem. 58, 539-547, 1975
 A; Title: Amino-acid sequence of the cyanogen-bromide fragment from human plasminogen that
 A; Reference number: A04626; MUID:76043692
 A; Molecule type: protein
 A; Residues: 483-507, 'E', 509-604 <W13>
 R; Robbins, K.C.; Bernabe, P.; Arzadon, L.; Summaria, L.; Biol. Chem. 248, 1631-1633, 1973
 A; Title: The primary structure of human Plasminogen. II. The histidine loop of human plasminogen
 A; Reference number: A92125; MUID:73149248
 A; Contents: annotation; active site
 R; Groskopf, W.R.; Summaria, L.; Robbins, K.C.; Biol. Chem. 244, 3590-3597, 1969
 A; Title: Studies on the active center of human plasmin. Partial amino acid sequence of a
 A; Reference number: A92048; MUID:69234739
 A; Contents: annotation; active site
 R; Prexier, M.; Viali, Z.; Patti, Y.; Biol. Chem. 257, 7401-7406, 1982
 A; Title: Structure of the omega-amino carboxylic acid-binding sites of human plasminogen.
 A; Reference number: A92382; MUID:82213905
 A; Contents: annotation; omega-amino carboxylic acid binding sites
 R; Viali, Z.; Patti, Y.; Biol. Chem. 259, 13690-13694, 1984
 A; Title: The fibrin-binding site of human plasminogen. Arginines 32 and 34 are essential
 A; Reference number: A92458; MUID:85054794
 A; Contents: annotation; fibrin binding site; omega-amino carboxylic acid binding site
 R; Cao, Y.; Ji, R.W.; Davidson, D.; Schallier, J.; Marti, D.; Soehndel, S.; McCance, S.G.; Biol. Chem. 271, 29461-29467, 1990
 A; Title: Kringle domains of human angiostatin. Characterization of the anti-proliferative
 A; Reference number: A58811; MUID:97067211
 A; Contents: annotation
 R; Liijnen, H.R.; Ugwu, F.; Bini, A.; Collen, D.; Biochemistry 37, 4699-4702, 1998
 A; Title: Generation of an angiostatin-like fragment from plasminogen by stromelysin-1 (M
 A; Reference number: A58812; MUID:9548733
 A; Contents: annotation
 R; Tulinsky, A.; Mulichak, A.M.; submitted to the Brookhaven Protein Data Bank, July 1991
 A; Reference number: A51341; PDB:1PK4
 A; Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 376-454
 R; Tulinsky, A.; Wu, T.P.; submitted to the Brookhaven Protein Data Bank, August 1993
 A; Reference number: A51911; PDB:1PKR
 A; Contents: annotation; X-ray crystallography, 2.48 angstroms, residues 102-181
 R; Padmanabhan, K.; Tulinsky, A.; submitted to the Brookhaven Protein Data Bank, April 1994
 A; Reference number: A52404; PDB:1PKM
 A; Contents: annotation; X-ray crystallography, 2.25 angstroms, residues 375-454
 R; Tulinsky, A.; Mathews, I.I.; submitted to the Brookhaven Protein Data Bank, December 1995
 A; Reference number: A52444; PDB:1CEA
 A; Title: The refined structure of the epsilon-amino caproic acid complex of human plasminogen
 A; Reference number: A8818; MUID:92031503
 A; Contents: annotation
 R; de Vos, A.M.; Uitsch, M.H.; Kelley, R.F.; Padmanabhan, K.; Tulinsky, A.; Westbrook, R.; Stec, B.; Teeter, M.M.; Whitlow, M.; Yamano, A.; submitted to the Brookhaven Protein Data Bank, June 1995
 A; Title: Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.4 angstroms
 A; Reference number: A35483; MUID:9218803
 A; Contents: annotation; X-ray crystallography, 2.4 angstroms
 R; Rejante, M.; Llinas, M.; submitted to the Brookhaven Protein Data Bank, August 1996
 A; Title: (1)H-NMR assignments and conformation by (1)H-NMR, residues 103-181
 A; Reference number: S43645; MUID:94237157
 A; Contents: annotation; conformation by (1)H-NMR
 R; Rejante, M.R.; Llinas, M.; Eur. J. Biochem. 221, 939-949, 1994
 A; Title: Solution structure of the epsilon-amino hexanoic acid complex of human plasminogen
 A; Reference number: A8817; MUID:94237158
 A; Contents: annotation; conformation by (1)H-NMR
 C; Comment: Plasminogen is synthesized by the kidney and is present in plasma and many
 C; Comment: plasminogen is converted to plasmin by plasminogen activators (see PIR:UKH
 C; Comment: plasmin is inactivated by alpha-2-antiplasmin (see PIR:ITHUA2) immediately
 C; Comment: without the inhibitor kringle 580, resulting in two chains connected by two disulfide bonds. Without
 C; Comment: microplasmin is formed by autolytic cleavage of plasmin under artificial con-
 C; Comment: stromelysin 1 (see PIR:KCHS1) acts on plasminogen to produce angiostatin.
 C; Genetics:
 A; Gene: GDB:PIG
 A; Cross references: GDB:119498; OMIM:173350
 A; Map position: 6q26-6q27
 C; Function:
 A; Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a
 A; Pathway: fibrinolysis
 C; Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homology
 C; Reference: angiogenesis inhibitor; blood; duplication; fibrinolysis; glycoprotein, homologous
 F; 1-19/Domain: plasminogen-related protein precursor homology <PLPH>
 F; 1-19/Domain: signal sequence #status predicted <SIG>
 F; 20-81/Domain: plasminogen #status experimental <PRO>
 F; 20-96/Domain: activation peptide
 F; 79-466/Domain: angiostatin #status experimental <APT>
 F; 97-580/Domain: plasmin #status experimental <MAT>
 F; 07-580/Domain: plasmin chain A #status experimental <CHA>
 F; 103-181/Domain: kringle homology <KR1>
 F; 185-262/Domain: kringle homology <KR2>
 F; 275-352/Domain: kringle homology <KR3>
 F; 377-454/Domain: kringle homology <KR4>
 F; 481-560/Domain: kringle homology <KR5>

F;550-580,581-810/Product: microplasmin #status experimental <MMT>

Query Match	20.6%	Score	504.5	DB	1	Length	810		
Best Local Similarity	41.1%	Pred. No.	2e-32						
Matches	108	Conservative	41	Mismatches	89	Indels	25	Gaps	8
Qy	203	OCTA---CCG---HRRGGYSSRIVGGNNMILLSOWPAASLQFO-GYHICGGSVITPLWII	253	Db	559	QCAAPSFDCGKPKQVEPKKCPGRVGGCVYAHPSWMPQVSILTRFGMHEGGTLISPEWVL	618		
Qy	254	TAACHCYVYLPLPKSWTIQVGL---VSULDNPAPSLVKEVYRHSKYPRLGNIALML	310	Db	619	TAACHCLEKSPRSPSYKVKIAGAHQEVNL----EHSVQELEVSFLEPLR--KQIAKKL	671		
Qy	311	AGPLTENENIQPVCLPNSEBNEPDGKVWISGMATED--GAGDASPVINHAAPLISK	368	Db	672	SSPAVITDKVPAICLPSNPWVADRTCECITGWGETQGFFGAG---LILKEAQLPVIENK	727		
Qy	369	ICNHRDVYGGITSPSMICAGYLUFGGVDSCQGDGSGGLYQERRLWKLVQATSGFGCAEV	428	Db	728	VCRNVEFLNGRVOSTELCAGHLAGTDSQCGDSSGGPLVCFEKDVKYILOCVTSGWILGCARP	787		
Qy	429	NKPGVYVTRVTSFLDWHQMDER	451	Db	788	NKPGVYVVRVSFVTVWIGVMRN	810		

Search completed: April 11, 2002, 08:53:02
Job time: 30 sec